

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2000, 23:47:05 ; Search time 556.28 Seconds
(without alignments)
-789.343 Million cell updates/sec

Title: US-09-099-898-1
Perfect score: 453
Sequence: 1 GTGTGGCGGATTGGTTAGC.....GGGAGCTGGAATAAACCT 453

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues
Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba1:
2: gb_ba2:
3: gb_om:
4: gb_ov:
5: gb_pat:
6: gb_ph:
7: gb_pi:
8: gb_pi2:
9: gb_pi1:
10: gb_pi2:
11: gb_pi3:
12: gb_pi4:
13: gb_pi5:
14: gb_pi6:
15: gb_pi7:
16: gb_pi8:
17: gb_pi9:
18: gb_pi10:
19: gb_pi11:
20: gb_pi12:
21: gb_pi13:
22: gb_pi14:
23: gb_pi15:
24: gb_pi16:
25: gb_pi17:
26: gb_pi18:
27: gb_pi19:
28: gb_pi20:
29: gb_pi21:
30: gb_pi22:
31: gb_pi23:
32: gb_pi24:
33: gb_pi25:
34: gb_pi26:
35: gb_pi27:
36: gb_pi28:
37: gb_pi29:
38: gb_pi30:
39: gb_pi31:
40: gb_pi32:
41: gb_pi33:
42: gb_pi34:
43: gb_pi35:
44: gb_pi36:

45: gb_htg7:
46: em_htg1:
47: em_htg2:
48: em_htg3:
49: em_hum5:
50: gb_pi13:
51: gb_pi14:
52: gb_pi15:
53: gb_pi16:
54: gb_pi17:
55: gb_pi18:
56: gb_pi19:
57: gb_pi20:
58: gb_pi21:

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
1	129	28.5	214575	45	AC008763	AC008763 Homo sapi
2	69.6	15.4	90442	57	AC016932	AC016932 Homo sapi
3	45.8	10.1	90442	57	AC016932	AC016932 Homo sapi
4	43.8	9.7	981	12	RNCRIMP	X54806 R. norvegicu
5	43.4	9.6	41055	1	SC684	AL033317 Streptomy
6	42.8	9.4	207651	55	AC009954	AC009954 Homo sapi
7	42.8	9.4	219565	53	AC009974	AC009974 Homo sapi
8	42.6	9.4	76968	54	AC022648	AC022648 Homo sapi
9	42.6	9.4	164520	44	AC020738	AC020738 Homo sapi
10	42.4	9.4	174707	43	AC017082	AC017082 Homo sapi
11	41.8	9.2	76199	55	AC023395	AC023395 Homo sapi
12	41.8	9.2	81213	55	AC023660	AC023660 Homo sapi
13	41.6	9.2	40883	11	AC005776	AC005776 Homo sapi
14	41.2	9.1	68523	55	AC023214	AC023214 Homo sapi
15	41.2	9.1	75135	55	AC005445	AC005445 Drosophila
16	40.8	9.0	5180	12	RN048596	U48596 Rattus norv
17	40.8	9.0	39441	11	AC005565	AC005565 Homo sapi
18	40.8	9.0	96079	56	AC023362	AC023362 Mus muscu
19	40.8	9.0	127778	44	AC020735	AC020735 Homo sapi
20	40.6	9.0	47852	1	MTV023	AL022022 Mycobacte
21	40.6	9.0	55482	54	AC022663	AC022663 Homo sapi
22	40.6	9.0	171480	44	AC021165	AC021165 Homo sapi
23	40.6	9.0	196696	55	AC005302	AC005302 Mus muscu
24	40.2	8.9	47852	1	MTV023	AL022022 Mycobacte
25	40.2	8.9	213455	32	AL135901	AL135901 Homo sapi
26	40.2	8.9	227070	56	AC023888	AC023888 Homo sapi
27	40	8.8	5880	16	H584	D14486 Equine herp
28	40	8.8	54775	55	AC023526	AC023526 Homo sapi
29	40	8.8	121882	44	AC019147	AC019147 Homo sapi
30	39.8	8.8	1320	3	CVCNYC	X95367 C.familiali
31	39.8	8.8	10167	2	SFU08223	U08223 Streptomyce
32	39.8	8.8	44300	54	AC022749	AC022749 Homo sapi
33	39.8	8.8	116789	54	AC015462	AC015462 Homo sapi
34	39.8	8.8	227178	55	AC023492	AC023492 Homo sapi
35	39.6	8.7	121882	44	AC019147	AC019147 Homo sapi
36	39.6	8.7	126650	44	AC019049	AC019049 Homo sapi
37	39.4	8.7	806	7	HV049482	U49482 Hordeum vul
38	39.4	8.7	2578	12	MUSHOXA	L08757 Mus musculu
39	39.4	8.7	51950	55	AC023561	AC023561 Homo sapi
40	39.4	8.7	127778	44	AC020735	AC020735 Homo sapi
41	39.2	8.7	2157	4	XELB1AA	M63663 X.borealis
42	39.2	8.7	3252	10	HSTAF113	Y11354 H.sapiens m
43	39.2	8.7	42189	11	AC005258	AC005258 Homo sapi
44	39.2	8.7	69798	54	AC022682	AC022682 Mus muscu
45	39.2	8.7	102713	44	AC019126	AC019126 Homo sapi

ALIGNMENTS

```
RESULT 1
AC008763
LOCUS
DEFINITION Homo sapiens chromosome 19 clone CTD-3214H19, LOW-PASS SEQUENCE
SAMPLING.
ACCESSION AC008763
VERSION AC008763.2 GI:6759002
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 214575)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 19
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 214575)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jan 26, 2000 this sequence version replaced gi:5686308.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
* NOTE: This record contains 143 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1
* 526: contig of 526 bp in length
* gap of unknown length
* 527 contig of 646 bp in length
* gap of unknown length
* 1173 1874: contig of 702 bp in length
* gap of unknown length
* 1875 2526: contig of 652 bp in length
* gap of unknown length
* 2527 3341: contig of 815 bp in length
* gap of unknown length
* 3342 3980: contig of 639 bp in length
* gap of unknown length
* 3981 4621: contig of 641 bp in length
* gap of unknown length
* 4622 5267: contig of 646 bp in length
* gap of unknown length
* 5268 5965: contig of 698 bp in length
* gap of unknown length
* 5966 6499: contig of 534 bp in length
* gap of unknown length
* 6500 6802: contig of 303 bp in length
* gap of unknown length
* 6803 7278: contig of 476 bp in length
* gap of unknown length
* 7279 7794: contig of 516 bp in length
* gap of unknown length
* 7795 8099: contig of 305 bp in length
* gap of unknown length
* 8100 8412: contig of 313 bp in length
* gap of unknown length
* 8413 8977: contig of 565 bp in length
* gap of unknown length
* 8978 9609: contig of 632 bp in length
* gap of unknown length
*
* 10398: contig of 789 bp in length
* gap of unknown length
* 11373: contig of 975 bp in length
* gap of unknown length
* 12245: contig of 872 bp in length
* gap of unknown length
* 12994: contig of 749 bp in length
* gap of unknown length
* 13817: contig of 823 bp in length
* gap of unknown length
* 14834: contig of 1017 bp in length
* gap of unknown length
* 15867: contig of 1033 bp in length
* gap of unknown length
* 16456: contig of 589 bp in length
* gap of unknown length
* 17173: contig of 717 bp in length
* gap of unknown length
* 18365: contig of 1192 bp in length
* gap of unknown length
* 19087: contig of 722 bp in length
* gap of unknown length
* 19955: contig of 868 bp in length
* gap of unknown length
* 20899: contig of 944 bp in length
* gap of unknown length
* 21726: contig of 827 bp in length
* gap of unknown length
* 22451: contig of 725 bp in length
* gap of unknown length
* 23825: contig of 1374 bp in length
* gap of unknown length
* 24745: contig of 920 bp in length
* gap of unknown length
* 25345: contig of 600 bp in length
* gap of unknown length
* 26371: contig of 1026 bp in length
* gap of unknown length
* 27436: contig of 1065 bp in length
* gap of unknown length
* 27957: contig of 521 bp in length
* gap of unknown length
* 28620: contig of 663 bp in length
* gap of unknown length
* 29584: contig of 964 bp in length
* gap of unknown length
* 30443: contig of 859 bp in length
* gap of unknown length
* 31147: contig of 704 bp in length
* gap of unknown length
* 32101: contig of 954 bp in length
* gap of unknown length
* 32865: contig of 764 bp in length
* gap of unknown length
* 33541: contig of 676 bp in length
* gap of unknown length
* 34142: contig of 601 bp in length
* gap of unknown length
* 35052: contig of 910 bp in length
* gap of unknown length
* 35799: contig of 747 bp in length
* gap of unknown length
* 36625: contig of 826 bp in length
* gap of unknown length
* 36835: contig of 210 bp in length
* gap of unknown length
* 37430: contig of 595 bp in length
* gap of unknown length
* 38409: contig of 979 bp in length
* gap of unknown length
* 38581: contig of 172 bp in length
* gap of unknown length
* 38656: contig of 75 bp in length
```


Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D.,
 Forcum-Tansey, J., Frantz, P., Ganesh, R., Gorrell, J.H., Gorrell, L.L.,
 Guevara, W., Harris, K., Hernandez, J., Hodgson, A., Hoques, M.,
 Holloway, C., Hosak, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, M.,
 Kelly, S., Kondesewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z.,
 Lichtarge, O., Liu, J., Liu, W., Logan, O., Lozano, R.J., Lu, J.,
 Luchter, R., Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M.,
 Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S.,
 Oswal, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L.,
 Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S.,
 Shah, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Sucgang, R.,
 Tabor, P., Taylor, T., Vasquez, L., Vinson, R., Vo, Q., Waibang, M.,
 Watling, K., Weinstock, G., Weinstock, I.R., Williamson, A.,
 Worley, K., Wren, J., Wrensford, G., Yu, W., Zhou, X., Nelson, D. and
 Gibbs, R.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT

Direct Submission
 Unpublished
 2 (bases 1 to 90442)

Worley, K.C.

Direct Submission

Submitted (09-DEC-1999) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Feb 19, 2000 this sequence version replaced gi:6552811.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: HMTH

Center clone name: Rp11-166C10

----- Summary Statistics

Sequencing vector: M13, L08821

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.980611

Consensus quality: 41098 bases at least Q40

Consensus quality: 53818 bases at least Q30

Consensus quality: 60396 bases at least Q20

Estimated insert size: 90442; agarose-fp estimation

Estimated insert size: 85446; sum-of-contigs estimation

Quality coverage: 0.7x in Q20 bases; agarose-fp estimation

Quality coverage: 0.7x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 53 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 811: contig of 811 bp in length

* 812 831: gap of unknown length

* 832 1876: contig of 1045 bp in length

* 1877 1896: gap of unknown length

* 1897 2882: contig of 986 bp in length

* 2883 2902: gap of unknown length

* 2903 3993: contig of 1091 bp in length

* 3994 4013: gap of unknown length

* 4014 4811: contig of 798 bp in length

* 4812 4831: gap of unknown length

* 4832 6227: contig of 1396 bp in length

* 6228 6247: gap of unknown length

* 6248 7283: contig of 1036 bp in length

* 7284 7303: gap of unknown length

* 7304 8543: contig of 1240 bp in length

* 8544 8563: gap of unknown length

* 8564 10252: contig of 1689 bp in length

* 10253 10272: gap of unknown length

* 10273 11441: contig of 1169 bp in length

* 11442 11461: gap of unknown length

* 11462 12487: contig of 1026 bp in length

12488 12507: gap of unknown length

12508 13621: contig of 1114 bp in length

13622 13641: gap of unknown length

13642 15080: contig of 1439 bp in length

15081 15100: gap of unknown length

15101 16137: contig of 1037 bp in length

16138 16157: gap of unknown length

16158 17466: contig of 1309 bp in length

17467 17486: gap of unknown length

17487 18999: contig of 1513 bp in length

18999 19019: gap of unknown length

19019 20438: contig of 1419 bp in length

20439 21658: gap of unknown length

21659 21678: gap of unknown length

21679 22207: contig of 529 bp in length

22208 22227: gap of unknown length

22228 23695: contig of 1468 bp in length

23696 23715: gap of unknown length

23716 24806: contig of 1091 bp in length

24807 24826: gap of unknown length

24827 26551: contig of 1725 bp in length

26552 26571: gap of unknown length

26572 27985: contig of 1394 bp in length

27986 29263: contig of 1278 bp in length

29264 29283: gap of unknown length

29284 30520: contig of 1237 bp in length

30521 30540: gap of unknown length

30541 32236: contig of 1696 bp in length

32237 32256: gap of unknown length

32257 33693: contig of 1437 bp in length

33694 33713: gap of unknown length

33714 35327: contig of 1614 bp in length

35328 35347: gap of unknown length

35348 36131: contig of 784 bp in length

36132 36151: gap of unknown length

36152 38604: contig of 2453 bp in length

38605 38624: gap of unknown length

38625 40511: contig of 1887 bp in length

40512 40531: gap of unknown length

40532 42147: contig of 1615 bp in length

42148 42166: gap of unknown length

42167 43749: contig of 1583 bp in length

43750 43769: gap of unknown length

43770 45010: contig of 1241 bp in length

45011 45030: gap of unknown length

45031 46545: contig of 1515 bp in length

46546 46565: gap of unknown length

46566 48528: contig of 1963 bp in length

48529 48548: gap of unknown length

48549 50692: contig of 2124 bp in length

50693 50692: gap of unknown length

50693 52614: contig of 1922 bp in length

52615 52634: gap of unknown length

52635 54090: contig of 1456 bp in length

54091 54110: gap of unknown length

54111 56624: contig of 2514 bp in length

56625 56644: gap of unknown length

56645 59368: contig of 2724 bp in length

59369 59388: gap of unknown length

59389 62256: contig of 2868 bp in length

62257 62276: gap of unknown length

62277 64634: contig of 2378 bp in length

64635 64674: gap of unknown length

64675 67214: contig of 2540 bp in length

67215 67234: gap of unknown length

67235 68685: contig of 1451 bp in length

68686 68705: gap of unknown length

68706 70178: contig of 1473 bp in length

70179 70198: gap of unknown length

70199 72544: contig of 2346 bp in length

72545 72564: gap of unknown length

```

JOURNAL      Unpublished
FEATURES     source
Location/Qualifiers
1. .981
/organism="Rattus norvegicus"
/strain="Wistar"
/db_xref="taxon:10116"
/dev_stage="1-2 day neonates"
/tissue_type="epidermis"
/cell_type="keratinocytes"
/clone_lib="plasmid, epidermal cDNA"
/clone="p.EL3.25"
<1..504
/codon_start=1
/product="cytokeratin type I"
/protein_id="CAA38577.1"
/db_xref="GI:55956"
/db_xref="SPTREMBL:O63115"
/translation="EPLDQVQIMNTVIAQMTATVVKTIKGFETEMSCFCTQDNVSV"

```

SOURCE ORGANISM
O-sialoglycoprotein endopeptidase; pantothenate kinase; phospho-sugar mutase; pseudouridylylate synthase; rplM; rpmJ; rpoA; rpsL; rpsK; rpsM; translational initiation factor IF1; trua; two-component regulator.
Streptomyces coelicolor A3(2).
Streptomyces coelicolor A3(2).

REFERENCE
Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 41055)
Saunders, D. and Harris, D.
Unpublished
2 (bases 1 to 41055)
Parkhill, J., Barrall, B.G. and Rajandream, M.A.
Direct Submission
Submitted (19-AUG-1998) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrall@sanger.ac.uk Cosmids supplied by Prof.
David A. Hrowood. I313b@tatescentre.norwich.ac.uk

Colney, Norwich, Norfolk NR4 7UH, UK
 3 (bases 1 to 41055)
 Redenbach, M., Kleser, H. M., Denapaite, D., Eichner, A., Cullum, J.,
 Kinashi, H., and Hopwood, D. A.
 A set of ordered cosmids and a detailed genetic and physical map
 for the 8 Mb Streptomyces coelicolor A3(2) chromosome
 Mol. Microbiol. 21 (1), 77-96 (1996)
 97000331
 NOTES:
 Streptomyces coelicolor sequencing at The Sanger Centre is funded
 by the BBSRC.
 Details of S. coelicolor sequencing at the Sanger Centre are
 available on the World Wide Web.
 (URL: <http://www.sanger.ac.uk/projects/S.coelicolor/>) CDS are
 numbered using the following system eg SC7B7.01c. SC (S.
 coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
 strand).
 The more significant matches with motifs in the PROSITE database
 are also included but some of these may be fortuitous. The length
 in codons is given for each CDS.
 Usually the highest scoring match found by fasta -o is given for
 CDS which show significant similarity to other CDS in the database.
 The position of possible ribosome binding site sequences are given
 where these have been used to deduce the initiation codon. Gene
 prediction is based on positional base preference in codons using a
 specially developed Hidden Markov Model (Krogh et al., Nucleic
 Acids Research, 22(22):4768-4778(1994)) and the Frameplot program
 of Bibb et al., Gene 30:157-66(1984) as implemented at
<http://www.nih.gov/jp/frameplot.pl>. CAUTION: We may not have predicted the
 correct initiation codon. Where possible we choose an initiation
 codon (atg, gtg, ttg or (att)) which is preceded by an upstream
 ribosome binding site sequence (Optimally 5-13bp before the
 initiation codon). If this cannot be identified we choose the most
 upstream initiation codon.
 IMPORTANT: This sequence MAY NOT be the entire insert of the
 sequenced clone. It may be shorter because we only sequence
 overlapping sections once, or longer, because we arrange for a
 small overlap between neighbouring submissions. Cosmid 6G4 lies
 between D31 and D63 in the AseI-D genomic restriction fragment.
 Location/Qualifiers
 1..41055
 /organism="Streptomyces coelicolor A3(2)"
 /strain="A3(2)"
 /db_xref="taxon:100226"
 /clone="cosmid 6G4"
 misc_feature
 1..1610
 /note="98.4% match to SCSECYDNA X83011 S.coelicolor secY
 locus DNA from 4547 to 6154"
 1..286
 /gene="adk"
 <1..286
 /gene="adk"
 /note="SC6G4.01, adk, adenylate kinase, partial CDS, len:
 >94 aa; almost identical to KAD_STRCO adenylate kinase (EC
 2.7.4.3) (205 aa) and highly similar to many e.g.
 KAD_BORPE adenylate kinase (EC 2.7.4.3) (218 aa), fasta
 scores; opt: 253 z-score: 429.1 E(): 1.2e-16, 50.6%
 identity in 85 aa overlap. Contains pfam match to entry
 PF00406 adenylatekinase, Adenylate kinases, score 136.80,
 E-value 3.2e-37"
 /codon_start=2
 /transl_table=11
 /product="adenylate kinase"
 /protein_id="CAA20379.1"
 /db_xref="GI:3449235"
 /db_xref="SPTREMBL:O86771"
 /translation="TAGRCVCEPEKHFVHTYTPPKKEGVCDCGGELYQRDDSDSE
 TVRKLEYHQTETFIIDYYSQGLVAIATGPDVETRRALKRKDQ"
 2..208
 /gene="adk"
 /note="pfam match to entry PF00406 adenylatekinase,
 Adenylate kinases, score 136.80, E-value 3.2e-37"
 misc_feature

RBS
 423..426
 /note="possible RBS upstream of map"
 gene
 438..1274
 /gene="map"
 CDS
 438..1274
 /gene="map"
 /note="SC6G4.02, map, methionine aminopeptidase, len: 278
 aa; identical to TR:O54208 (EMBL:X83011) methionine
 aminopeptidase (278 aa) and highly similar to many
 e.g. AMPM_PACSU methionine aminopeptidase (EC 3.4.11.18)
 (248 aa), fasta scores; opt: 546 z-score: 922.2 E(): 0,
 44.5% identity in 256 aa overlap. Also similar to S.
 coelicolor map2 (E): 3.2e-33, 46.8% identity in 267 aa
 overlap). Contains PS00680 Methionine aminopeptidase
 subfamily 1 signature and pfam match to entry PF00557
 pep_M24, metalloproteinase family M24, score 194.80,
 E-value 1.4e-54"
 /codon_start=1
 /transl_table=11
 /product="methionine aminopeptidase"
 /protein_id="CAA20380.1"
 /db_xref="GI:3449236"
 /db_xref="SPTREMBL:O54208"
 /translation="MVQIKNPEQIAKMEAGLVVAAIHAATREAAVPGATKDLDOVA
 RYLAEDHAKPNFLGSGFPFATICTSVNEVVVHGIPSDDVVKDGVISIDCGALIDG
 WHGDAATAFVSGHSPFLVELSKVTEESWAGIAMKQGNLRKLVDSRAIDETIRROP
 KPGGKYGIIDYGGHIGTEMHMDPHLLNYVRRRGKPKLPFGCLAIEPMVSLGT
 TRTEVLPDENTVITTDGTSSSHWHSVALTEQGPLVLTSPPDGKAKLAELGITAAPDP
 LA"
 misc_feature
 444..1196
 /gene="map"
 /note="pfam match to entry PF00557 pep_M24,
 metalloproteinase family M24, score 194.80, E-value
 1.4e-54"
 954..1010
 /gene="map"
 /note="PS00680 Methionine aminopeptidase subfamily 1
 signature"
 1405..1408
 /note="possible RBS upstream of infA"
 gene
 1416..1637
 /gene="infA"
 1416..1637
 /gene="infA"
 /note="SC6G4.03, infA, translational initiation factor
 IF1, len: 73 aa; identical to TR:O54209 (EMBL:X83011)
 translational initiation factor IF1 (fragment) (65 aa) and
 highly similar to many e.g. IF1_PACSU translation
 initiation factor IF-1 (71 aa), fasta scores; opt: 384
 z-score: 792.5 E(): 0, 78.6% identity in 70 aa overlap.
 Contains pfam match to entry PF00575 S1, S1 RNA binding
 motif, score 54.10, E-value 3.1e-12"
 /codon_start=1
 /transl_table=11
 /product="translational initiation factor IF1"
 /protein_id="CAA20381.1"
 /db_xref="GI:3449237"
 /translation="MAKQGAIEGTVEESLPNAMEKVELQNGHVLAHISGKMRMH
 YKILPDDRVRVVELSPYDLTRGRIVRYK"
 1422..1634
 /gene="infA"
 /note="pfam match to entry PF00575 S1, S1 RNA binding
 motif, score 54.10, E-value 3.1e-12"
 RBS
 1690..1693
 /note="possible RBS upstream of rpmJ"
 misc_feature
 1706..1816
 /gene="rpmJ"
 /note="pfam match to entry PF00444 L36, Ribosomal protein
 L36, score 70.40, E-value 3.1e-22"
 gene
 1706..1819
 /gene="rpmJ"
 CDS
 1706..1819
 /gene="rpmJ"

[illegible]


```

* * * gap of unknown length
* * * 45715 46634: contig of 920 bp in length
* * * 46635 47535: contig of 901 bp in length
* * * 47536 48461: contig of 926 bp in length
* * * 48462 49352: contig of 891 bp in length
* * * 49353 50255: contig of 903 bp in length
* * * 50256 51180: contig of 925 bp in length
* * * 51181 52074: contig of 894 bp in length
* * * 52075 52993: contig of 919 bp in length
* * * 52994 53936: contig of 943 bp in length
* * * 53937 54853: contig of 917 bp in length
* * * 54854 55782: contig of 929 bp in length
* * * 55783 56714: contig of 932 bp in length
* * * 56715 57619: contig of 905 bp in length
* * * 57620 58556: contig of 937 bp in length
* * * 58557 59455: contig of 899 bp in length
* * * 59456 60384: contig of 929 bp in length
* * * 60385 61280: contig of 896 bp in length
* * * 61281 62186: contig of 906 bp in length
* * * 62187 63081: contig of 895 bp in length
* * * 63082 63964: contig of 883 bp in length
* * * 63965 64876: contig of 912 bp in length
* * * 64877 65831: contig of 955 bp in length

Query Match
Best Local Similarity 9.4%; Score 42.6; DB 54; Length 76968;
Matches 122; Conservative 0; Mismatches 112; Indels 4; Gaps 1;

QY 210 CTTCCAGGGGACCTGGTACTTGCCTCCCGAGGCTT---CGCGTCACCGGTCGCACT 265
DB 24356 CGCCCGCGGGCCCGCGCCCGCCCGCGGGGCGGGGCGGGGCGGGGCGGGGCGG 24415

QY 266 TGTGGTCCGCTGTGCTGTGGATGTGCGGCGCGGAGACCATGTCATCTCCAGTGC 325
DB 24416 GGGCCCCCGGGGGGGCGCGCGCGCGCGGGGGGGCGCGCGGGGCGGGCGG 24475

QY 326 GGGGGATGACATGACGACGAGCGCGTGTCTGTGTCAGGCTCGAGTGGCGGCAG 385
DB 24476 CGCGCGCGCGCGCGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 24535

QY 386 CGGTGTCACAGCGCGGCGGAGCGGCTCCAGTTCGAGGGGTGCGGGGAGCTGG 443
DB 24536 CGCGCGCGCGCGCGGGGCGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGG 24593

```

RESULT 9
 AC020738
 LOCUS AC020738 164520 bp DNA HTG 08-JAN-2000
 DEFINITION Homo sapiens clone RP11-753114, LOW-PASS SEQUENCE SAMPLING.
 ACCESSION AC020738
 VERSION AC020738.1 GI:6682695
 KEYWORDS HTG; HTGS_PHASE0.

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 164520)
 AUTHORS Waterston,R.H.
 TITLE The sequence of Homo sapiens clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 164520)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (08-JAN-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 COMMENT
 * Center Project name: H_NH0753114.
 * NOTE: This record contains 231 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.
 * 1
 * 706: contig of 706 bp in length
 * 707: gap of unknown length
 * 716: contig of 518 bp in length
 * 717: gap of unknown length
 * 1234: contig of 733 bp in length
 * 1235: gap of unknown length
 * 1245: contig of 731 bp in length
 * 1978: gap of unknown length
 * 1988: contig of 525 bp in length
 * 2719: gap of unknown length
 * 2729: contig of 517 bp in length
 * 3254: gap of unknown length
 * 3264: contig of 714 bp in length
 * 3781: gap of unknown length
 * 3791: contig of 710 bp in length
 * 4505: gap of unknown length
 * 4515: contig of 500 bp in length
 * 5225: gap of unknown length
 * 5235: contig of 730 bp in length
 * 5735: gap of unknown length
 * 5745: contig of 707 bp in length
 * 6475: gap of unknown length
 * 6485: contig of 728 bp in length
 * 6485: gap of unknown length
 * 7191: contig of 728 bp in length
 * 7192: gap of unknown length
 * 7202: contig of 728 bp in length
 * 7930: gap of unknown length
 * 7940: contig of 728 bp in length
 * 8668: gap of unknown length
 * 8678: contig of 247 bp in length
 * 8925: gap of unknown length
 * 8935: contig of 723 bp in length
 * 9658: gap of unknown length
 * 9668: contig of 518 bp in length
 * 10186: gap of unknown length
 * 10196: contig of 516 bp in length
 * 10712: gap of unknown length
 * 10722: contig of 515 bp in length
 * 11237: gap of unknown length
 * 11247: contig of 513 bp in length
 * 11760: gap of unknown length
 * 11770: contig of 514 bp in length
 * 12284: gap of unknown length
 * 12294: contig of 323 bp in length
 * 12617: gap of unknown length
 * 12627: contig of 724 bp in length
 * 13351: gap of unknown length
 * 13361: contig of 517 bp in length
 * 13878: gap of unknown length
 * 13888: contig of 515 bp in length

* 14403 14412: gap of unknown length
* 14413 15117: contig of 705 bp in length
* 15118 15127: gap of unknown length
* 15128 16053: contig of 926 bp in length
* 16054 16063: gap of unknown length
* 16064 17174: contig of 1111 bp in length
* 17175 17184: gap of unknown length
* 17185 17917: contig of 733 bp in length
* 17918 17927: gap of unknown length
* 17928 18667: contig of 740 bp in length
* 18668 18677: gap of unknown length
* 18678 19426: contig of 749 bp in length
* 19436 19436: gap of unknown length
* 19437 20191: contig of 755 bp in length
* 20192 20201: gap of unknown length
* 20202 20926: contig of 725 bp in length
* 20927 20936: gap of unknown length
* 20937 21653: contig of 719 bp in length
* 21654 21656: gap of unknown length
* 21656 22400: contig of 735 bp in length
* 22401 22410: gap of unknown length
* 22411 22923: contig of 515 bp in length
* 22924 22933: gap of unknown length
* 22934 23450: contig of 515 bp in length
* 23451 23460: gap of unknown length
* 23461 23972: contig of 512 bp in length
* 23973 23982: gap of unknown length
* 23983 24722: contig of 740 bp in length
* 24723 24732: gap of unknown length
* 24733 25060: contig of 318 bp in length
* 25061 25060: gap of unknown length
* 25061 25765: contig of 705 bp in length
* 25766 25775: gap of unknown length
* 25776 26511: contig of 738 bp in length
* 26512 26523: gap of unknown length
* 26524 27260: contig of 737 bp in length
* 27261 27270: gap of unknown length
* 27271 27793: contig of 523 bp in length
* 27794 27803: gap of unknown length
* 27804 28538: contig of 735 bp in length
* 28539 28548: gap of unknown length
* 28549 29062: contig of 514 bp in length
* 29063 29072: gap of unknown length
* 29073 29585: contig of 513 bp in length
* 29586 29595: gap of unknown length
* 29596 30108: contig of 513 bp in length
* 30109 30634: contig of 516 bp in length
* 30635 30644: gap of unknown length
* 30645 31385: contig of 741 bp in length
* 31386 31395: gap of unknown length
* 31396 32384: contig of 989 bp in length
* 32385 32394: gap of unknown length
* 32395 33144: contig of 750 bp in length
* 33145 33154: gap of unknown length
* 33155 33891: contig of 737 bp in length
* 33892 33901: gap of unknown length
* 33902 34637: contig of 736 bp in length
* 34638 34648: gap of unknown length
* 34649 35160: contig of 513 bp in length
* 35161 35170: gap of unknown length
* 35171 35685: contig of 515 bp in length
* 35686 35695: gap of unknown length
* 35696 36127: contig of 432 bp in length
* 36128 36137: gap of unknown length
* 36138 36651: contig of 514 bp in length
* 36652 36661: gap of unknown length
* 36662 37395: contig of 734 bp in length
* 37396 37405: gap of unknown length
* 37406 38129: contig of 724 bp in length
* 38130 38139: gap of unknown length
* 38140 38593: contig of 454 bp in length
* 38594 38603: gap of unknown length

* 38604 39023: contig of 420 bp in length
* 39024 39033: gap of unknown length
* 39034 39549: contig of 516 bp in length
* 39550 39559: gap of unknown length
* 39560 40309: contig of 750 bp in length
* 40310 40319: gap of unknown length
* 40320 41044: contig of 725 bp in length
* 41045 41054: gap of unknown length
* 41055 41572: contig of 518 bp in length
* 41573 41582: gap of unknown length
* 41583 42095: contig of 513 bp in length
* 42096 42105: gap of unknown length
* 42106 42619: contig of 514 bp in length
* 42620 42629: gap of unknown length
* 42630 43143: contig of 514 bp in length
* 43144 43153: gap of unknown length
* 43154 43651: contig of 498 bp in length
* 43652 43661: gap of unknown length
* 43662 44171: contig of 510 bp in length
* 44172 44181: gap of unknown length
* 44182 44918: contig of 737 bp in length
* 44919 44928: gap of unknown length
* 44929 45389: contig of 461 bp in length
* 45390 45399: gap of unknown length
* 45400 45917: contig of 518 bp in length
* 45918 45927: gap of unknown length
* 45928 46444: contig of 517 bp in length
* 46445 46454: gap of unknown length
* 46455 46875: contig of 421 bp in length
* 46876 46885: gap of unknown length
* 46886 47331: contig of 446 bp in length
* 47332 47341: gap of unknown length
* 47342 47735: contig of 394 bp in length
* 47736 47745: gap of unknown length
* 47746 48373: contig of 628 bp in length
* 48374 48383: gap of unknown length
* 48384 49073: contig of 690 bp in length
* 49074 49083: gap of unknown length
* 49084 49820: contig of 737 bp in length
* 49821 49830: gap of unknown length
* 49831 50566: contig of 736 bp in length
* 50567 50576: gap of unknown length
* 50577 51319: contig of 743 bp in length
* 51320 51329: gap of unknown length
* 51330 51845: contig of 516 bp in length
* 51846 51855: gap of unknown length
* 51856 52366: contig of 511 bp in length
* 52367 52377: gap of unknown length
* 52377 52890: contig of 514 bp in length

Query Match 9.4%; Score 42.6; DB 44; Length 164520;
Best Local Similarity 50.6%; Pred. No. 3.4;
Matches 128; Conservative 0; Mismatches 124; Indels 1; Gaps 1;

Qy 188 GGCCTGGAGTGCAGAGCGTCACTCCAGGGGGGACCTGGCTACTTGCCTCCCGAGGCTTC 247
Db 120880 GCGCGCTGCGGGAGGGGGGCGCCCGGGGGAGAGGGCGGGTGGCCCGCAGGCACA 120939
Qy 248 GCGGTACCGGTGCACCTTGTGGCTCCCGCTCGCTGGCTGGGTATGTGCGGCCGAGACC 307
Db 120940 GCGGTGCGCCCGCGCGTGTGCGAGTGTGTGCGGGTGGCGCGCGCGGCC 120999
Qy 308 ACATGCTACTGCCAGTGGCGGGCATGACTGGACGGAGCGGCTGCTGCTGTGCTGCA 366
Db 121000 CCGCGCGGTGCGCGTCCGCTCGCGCGGCTGAGGGGGGCGCGCGCGGGGG 121059
Qy 367 GCGCTGAGTGCAGCGCGTGCACAGCGCGGCGAGCGGCTCCAGGTCCGAGG 426
Db 121060 GCGGGGTGAGTGGCGGGGCGCTGGCGGCTGTGTTGGCGGTGTGCGGGCGCTG 121119
Qy 427 GGTGCGGGGAG 439
Db 121120 GCGCGGGCGGG 121132

```

RESULT 10
AC017082
LOCUS AC017082 174707 bp DNA HTG 09-DEC-1999
DEFINITION Homo sapiens clone RP11-472M4, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC017082
VERSION AC017082.1 GI:6554065
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 174707)
Waterston.R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 174707)
Waterston.R.H.
Direct Submission
Submitted (09-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
Center project name: H_NH0567M19.
* NOTE: This record contains 326 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
386: contig of 386 bp in length
387
396: gap of unknown length
397
910: contig of 514 bp in length
911
920: gap of unknown length
921
1200: contig of 280 bp in length
1201
1210: gap of unknown length
1211
1722: contig of 512 bp in length
1723
1732: gap of unknown length
1733
2078: contig of 346 bp in length
2079
2088: gap of unknown length
2089
2800: contig of 512 bp in length
2801
2810: gap of unknown length
2811
3122: contig of 512 bp in length
3123
3132: gap of unknown length
3133
3643: contig of 511 bp in length
3644
3653: gap of unknown length
3654
4166: contig of 513 bp in length
4167
4176: gap of unknown length
4177
4689: contig of 513 bp in length
4690
4699: gap of unknown length
4700
5211: contig of 512 bp in length
5212
5221: gap of unknown length
5222
5666: contig of 445 bp in length
5667
5676: gap of unknown length
5677
6190: contig of 514 bp in length
6191
6200: gap of unknown length
6201
6714: contig of 514 bp in length
6715
7244: gap of unknown length
7236: contig of 512 bp in length
7237
7246: gap of unknown length
7247
7729: contig of 483 bp in length
7730
7739: gap of unknown length
7740
8591: contig of 852 bp in length
8592
8601: gap of unknown length
8602
9113: contig of 512 bp in length
9114
9123: gap of unknown length
9124
9634: contig of 511 bp in length
9635
9644: gap of unknown length

10156: contig of 512 bp in length
10166: gap of unknown length
10167
10670: contig of 504 bp in length
10680: gap of unknown length
10681
10917: contig of 237 bp in length
10918
10927: gap of unknown length
10928
11132: contig of 205 bp in length
11133
11142: gap of unknown length
11143
11648: contig of 506 bp in length
11649
11658: gap of unknown length
11659
12170: contig of 512 bp in length
12171
12180: gap of unknown length
12181
13038: contig of 858 bp in length
13039
13048: gap of unknown length
13049
13472: contig of 424 bp in length
13473
13482: gap of unknown length
13483
13994: contig of 512 bp in length
13995
14004: gap of unknown length
14005
14436: contig of 432 bp in length
14437
14446: gap of unknown length
14447
14955: contig of 509 bp in length
14956
14965: gap of unknown length
14966
15477: contig of 512 bp in length
15478
15487: gap of unknown length
15488
15999: contig of 512 bp in length
16000
16009: gap of unknown length
16010
16397: contig of 388 bp in length
16398
16407: gap of unknown length
16408
16863: contig of 456 bp in length
16864
16873: gap of unknown length
16874
17385: contig of 512 bp in length
17386
17395: gap of unknown length
17396
18220: contig of 825 bp in length
18221
18230: gap of unknown length
18231
18742: contig of 512 bp in length
18743
18752: gap of unknown length
18753
18876: contig of 124 bp in length
18877
18886: gap of unknown length
18887
19400: contig of 514 bp in length
19401
19410: gap of unknown length
19411
19573: contig of 183 bp in length
19574
19583: gap of unknown length
19584
20423: contig of 846 bp in length
20430
20439: gap of unknown length
20440
21291: contig of 852 bp in length
21301
21301: gap of unknown length
21302
21775: contig of 474 bp in length
21785: gap of unknown length
21786
22296: contig of 511 bp in length
22297
22306: gap of unknown length
22307
22810: contig of 504 bp in length
22811
22820: gap of unknown length
22821
23332: contig of 512 bp in length
23342: gap of unknown length
23343
23854: contig of 512 bp in length
23855
23864: gap of unknown length
23865
24376: contig of 512 bp in length
24377
24386: gap of unknown length
24387
24900: contig of 514 bp in length
24910: gap of unknown length
24911
25424: contig of 514 bp in length
25425
25434: gap of unknown length
25428: contig of 394 bp in length
25435
25838: gap of unknown length
25829
26350: contig of 512 bp in length
25839
26360: gap of unknown length
26351
26361: contig of 511 bp in length
26361
26871: contig of 511 bp in length
26872
26881: gap of unknown length
26882
28779: contig of 1898 bp in length
28780
28789: gap of unknown length
28790
29300: contig of 511 bp in length
29301
29310: gap of unknown length
29311
29774: contig of 464 bp in length

```

COMMENT

		29775	29784:	gap of unknown length	29775	29784:	Score 42.4;	DB 43;	Length 174707;
*		29785	30721:	contig of 943 bp in length	29785	30721:	Best Local Similarity 48.7%;	pred. No. 3,7;	
*		30737:	gap of unknown length		30737:	gap of unknown length	Matches 115;	Conservative	
*		30738	31251:	contig of 514 bp in length	30738	31251:	0;	Mismatches 121;	Indels 0; Gaps
*		31252:	gap of unknown length		31252:	gap of unknown length			
*		31262	31773:	contig of 512 bp in length	31262	31773:			
*		31774	31783:	gap of unknown length	31774	31783:			
*		31784	32199:	contig of 416 bp in length	31784	32199:			
*		32200	32209:	gap of unknown length	32200	32209:			
*		32210	32721:	contig of 512 bp in length	32210	32721:			
*		32722	32731:	gap of unknown length	32722	32731:			
*		32732	32977:	contig of 246 bp in length	32732	32977:			
*		32978	32987:	gap of unknown length	32978	32987:			
*		32988	33467:	contig of 480 bp in length	32988	33467:			
*		33468	33477:	gap of unknown length	33468	33477:			
*		33478	33991:	contig of 514 bp in length	33478	33991:			
*		33992	34001:	gap of unknown length	33992	34001:			
*		34002	34514:	contig of 513 bp in length	34002	34514:			
*		34515	34524:	gap of unknown length	34515	34524:			
*		34525	35373:	contig of 849 bp in length	34525	35373:			
*		35384	35383:	gap of unknown length	35384	35383:			
*		35385	35895:	contig of 512 bp in length	35385	35895:			
*		35896	35905:	gap of unknown length	35896	35905:			
*		35906	36339:	contig of 434 bp in length	35906	36339:			
*		36340	36349:	gap of unknown length	36340	36349:			
*		36350	37213:	contig of 854 bp in length	36350	37213:			
*		37214	37217:	contig of 214 bp in length	37214	37217:			
*		37428	37437:	gap of unknown length	37428	37437:			
*		37438	37949:	contig of 512 bp in length	37438	37949:			
*		37950	37959:	gap of unknown length	37950	37959:			
*		37960	38829:	contig of 870 bp in length	37960	38829:			
*		38830	38839:	gap of unknown length	38830	38839:			
*		38840	39343:	contig of 504 bp in length	38840	39343:			
*		39344	39353:	gap of unknown length	39344	39353:			
*		39354	39670:	contig of 317 bp in length	39354	39670:			
*		39671	39680:	gap of unknown length	39671	39680:			
*		39681	40193:	contig of 513 bp in length	39681	40193:			
*		40194	40203:	gap of unknown length	40194	40203:			
*		40204	40524:	contig of 321 bp in length	40204	40524:			
*		40525	40534:	gap of unknown length	40525	40534:			
*		40535	41048:	contig of 514 bp in length	40535	41048:			
*		41049	41058:	gap of unknown length	41049	41058:			
*		41059	41141:	contig of 83 bp in length	41059	41141:			
*		41142	41151:	gap of unknown length	41142	41151:			
*		41152	41664:	contig of 513 bp in length	41152	41664:			
*		41665	41674:	gap of unknown length	41665	41674:			
*		41675	42171:	contig of 497 bp in length	41675	42171:			
*		42172	42181:	gap of unknown length	42172	42181:			
*		42182	42695:	contig of 514 bp in length	42182	42695:			
*		42696	42705:	gap of unknown length	42696	42705:			
*		42706	43219:	contig of 514 bp in length	42706	43219:			
*		43220	43229:	gap of unknown length	43220				

repeat_region	1891. .2012	/rpt_family="AluJb"	18393. .18690	/rpt_family="AluSc"
repeat_region	2024. .2326	/rpt_family="AluY"	18706. .18968	
repeat_region	2327. .2616	/rpt_family="AluSx"	/note="BLASTN similarity to T41380 (105. .368); match: 0.99, score: 3.8e-100; database searched: est; phib12_19/1TV Homo sapiens CDNA clone phib12_19/1TV."	
repeat_region	2776. .2841	/rpt_family="LINE2"	19418. .19703	
repeat_region	3495. .3622	/rpt_family="LINE2"	/rpt_family="AluSx"	
repeat_region	3667. .3766	/rpt_family="MER3"	complement(20667. .20801)	
misc_feature	complement(4287. .4396)	/note="predicted exon, program: grail2exons_human_1.3, frame: 2, quality: good, score: 52.000"		
repeat_region	4507. .4787	/rpt_family="AluSx"	complement(20802. .20829)	
repeat_region	4808. .5109	/rpt_family="AluY"	complement(20871. .20955)	
repeat_region	5283. .5545	/rpt_family="AluSx"	complement(20982. .21134)	
repeat_region	6064. .6207	/rpt_family="MIR"	complement(21237. .21352)	
repeat_region	6537. .6577	/rpt_family="(TGG)n"	complement(21349. .21470)	
repeat_region	8122. .8401	/rpt_family="AluJb"	21482. .21559	
repeat_region	8402. .8565	/rpt_family="AluJo/FRAM"	complement(22010. .22285)	
repeat_region	8585. .8889	/rpt_family="AluJb"	22319. .22464	
repeat_region	9886. .10177	/rpt_family="AluSx"	22465. .22757	
repeat_region	10179. .10251	/rpt_family="(GAA)n"	22758. .22935	
repeat_region	complement(11205. .11323)	/rpt_family="AluSx"	complement(23192. .23484)	
repeat_region	11363. .11494	/rpt_family="AluY"	23530. .23829	
repeat_region	11506. .11820	/rpt_family="AluY"	complement(24456. .24584)	
repeat_region	11821. .12003	/rpt_family="AluY"	24597. .24734	
repeat_region	12022. .12128	/rpt_family="FLAM_A"	24737. .25071	
repeat_region	complement(12200. .12259)	/rpt_family="AluSx"	25072. .25238	
repeat_region	complement(12910. .13002)	/rpt_family="MIR"	complement(25239. .25317)	
misc_feature	13871. .14036	/note="predicted exon, program: grail2exons_human_1.3, frame: 1, quality: excellent, score: 83.000"		
repeat_region	14662. .14951	/rpt_family="MIR"	complement(26347. .26643)	
repeat_region	complement(15085. .15108)	/rpt_family="AluSp"	complement(26647. .26831)	
repeat_region	15109. .15431	/rpt_family="AluSx"	complement(26839. .27000)	
misc_feature	15662. .15770	/note="predicted exon, program: grail2exons_human_1.3, frame: 1, quality: excellent, score: 80.000"		
repeat_region	16107. .16190	/rpt_family="LMB8"	complement(27094. .27394)	
repeat_region	16191. .16497	/rpt_family="AluSx"	complement(27415. .27507)	
repeat_region	16498. .16937	/rpt_family="LMB8"	27599. .27892	
repeat_region	complement(17602. .17708)	/rpt_family="LINE2"	27924. .28121	
repeat_region	complement(17709. .18005)	/rpt_family="AluSp"	/note="BLASTX similarity to 619788 (14. .79); match: 0.5, score: 6.7e-17; database searched: nr; (L38826) L21 ribosomal protein [Homo sapiens]"	
repeat_region			28037. .28147	
repeat_region			/note="BLASTX similarity to 619788 (52. .88); match: 0.81, score: 7.0e-15; database searched: nr; (L38826) L21 ribosomal protein [Homo sapiens]"	
repeat_region			28159. .28393	
Query Match 9.2%; Score 41.6; DB 11; Length 40883;				
Best Local Similarity 45.2%; Pred. No. 6.6;				
Matches 152; Conservative 0; Mismatches 184; Indels 0; Gaps 0;				

```
QY 108 TGTGCTCCATGAAGACCATCAATGAGAGGATCCAGAGGTCGCGGCTCCCTAATAT 167
  || || || || || || || || || || || || || || || || || || || || ||
Db 21212 TGCCTTTGGGGGGGGGACGACACTGGCGGCTGCGGAGGCTCGCGGCGTCCGCG 21153
  || || || || || || || || || || || || || || || || || || || || ||
QY 168 TTAGGGCAATAGACCAATGGCTGGAGTGCAGAGCGTCACCTCCAGGGGGGACTGG 227
  || || || || || || || || || || || || || || || || || || || || ||
Db 21152 TAACTGATTCGTAATGGGCGGGGGCGGCCACCTACGCGATCCGCGGCGCGGG 21093
  || || || || || || || || || || || || || || || || || || || || ||
QY 228 CTACTTGCCCGGAGCTTCGCGGTACCGGTGCACTTGTGCTCGCGCTGGCTGCTCGT 287
  || || || || || || || || || || || || || || || || || || || || ||
Db 21092 GCGGGGAGCGGGGCTACGCGGGCTCGGGGCTCGGGGCTCCGCGCGCTGCTCG 21033
  || || || || || || || || || || || || || || || || || || || || ||
QY 288 GGGATGTGCGCGCCAGACCATGTCACCTGCGGAGTGCAGGCGGCGGACTGGACCGGAG 347
  || || || || || || || || || || || || || || || || || || || || ||
Db 21032 GCGGGGGGCGCTGCGCGTGTGGTCTGGGGCGGGGCTGCTCGGCTCTCTGCC 20973
  || || || || || || || || || || || || || || || || || || || || ||
QY 348 CGCGTGTGCTGTGTCAGCCCTGAGTGCAGCGCGCGGCGGCGGTCACAGCGGGGGAG 407
  || || || || || || || || || || || || || || || || || || || || ||
Db 20972 TTTGCTGCGCTGAGACCGCGGCTTGGCTGCGCGGGGGCGGGTGGGGGCGCGGGC 20913
  || || || || || || || || || || || || || || || || || || || || ||
QY 408 GCGGCTCCAGGTCGAGGGTTGCGGGGAGCTGG 443
  || || || || || || || || || || || || || || || || || || || || ||
Db 20912 GCGCGGGGCGGGCGCGCTCGGGGCGGGGCGGGG 20877
  || || || || || || || || || || || || || || || || || || || || ||

RESULT 14
AC023214/C 68523 bp DNA HTG 09-FEB-2000
LOCUS Homo sapiens clone RP11-313B12, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC023214
ACCESSION AC023214
VERSION AC023214.1 GI:6957756
KEYWORDS HTG; HTGS-PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 68523)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-313B12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 68523)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Chapel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferrelira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
DIRECT SUBMISSION
Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6599
Center clone name: 313_B12
```

```
-----
* NOTE: This record contains 88 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1
* 786: contig of 786 bp in length
* gap of unknown length
* 787 contig of 763 bp in length
* gap of unknown length
* 1550 contig of 785 bp in length
* gap of unknown length
* 2335 contig of 770 bp in length
* gap of unknown length
* 3105 contig of 776 bp in length
* gap of unknown length
* 3881 contig of 772 bp in length
* gap of unknown length
* 4653 contig of 770 bp in length
* gap of unknown length
* 5423 contig of 766 bp in length
* gap of unknown length
* 6189 contig of 781 bp in length
* gap of unknown length
* 6970 contig of 769 bp in length
* gap of unknown length
* 7739 contig of 768 bp in length
* gap of unknown length
* 8507 contig of 754 bp in length
* gap of unknown length
* 9261 contig of 784 bp in length
* gap of unknown length
* 10045 contig of 782 bp in length
* gap of unknown length
* 10827 contig of 781 bp in length
* gap of unknown length
* 11608 contig of 793 bp in length
* gap of unknown length
* 12401 contig of 766 bp in length
* gap of unknown length
* 13167 contig of 768 bp in length
* gap of unknown length
* 13935 contig of 773 bp in length
* gap of unknown length
* 14708 contig of 763 bp in length
* gap of unknown length
* 15471 contig of 860 bp in length
* gap of unknown length
* 16331 contig of 772 bp in length
* gap of unknown length
* 17103 contig of 763 bp in length
* gap of unknown length
* 17866 contig of 780 bp in length
* gap of unknown length
* 18646 contig of 760 bp in length
* gap of unknown length
* 19406 contig of 797 bp in length
* gap of unknown length
* 20203 contig of 797 bp in length
* gap of unknown length
* 21000 contig of 779 bp in length
* gap of unknown length
* 21779 contig of 768 bp in length
* gap of unknown length
* 22547 contig of 772 bp in length
* gap of unknown length
* 23319 contig of 788 bp in length
* gap of unknown length
```

*	24107	24896:	gap of 790 bp in length
*			contig of 790 bp in length
*	24897	25670:	gap of unknown length
*			contig of 774 bp in length
*	25671	26434:	gap of unknown length
*			contig of 764 bp in length
*	26435	27198:	gap of unknown length
*			contig of 764 bp in length
*	27199	27981:	gap of unknown length
*			contig of 783 bp in length
*	27982	28766:	gap of unknown length
*			contig of 785 bp in length
*	28767	29562:	gap of unknown length
*			contig of 796 bp in length
*	29563	30377:	gap of unknown length
*			contig of 815 bp in length
*	30378	31157:	gap of unknown length
*			contig of 780 bp in length
*	31158	31937:	gap of unknown length
*			contig of 780 bp in length
*	31938	32717:	gap of unknown length
*			contig of 780 bp in length
*	32718	33476:	gap of unknown length
*			contig of 759 bp in length
*	33477	34237:	gap of unknown length
*			contig of 761 bp in length
*	34238	34999:	gap of unknown length
*			contig of 762 bp in length
*	35000	35755:	gap of unknown length
*			contig of 756 bp in length
*	35756	36513:	gap of unknown length
*			contig of 758 bp in length
*	36514	37262:	gap of unknown length
*			contig of 749 bp in length
*	37263	38033:	gap of unknown length
*			contig of 771 bp in length
*	38034	38803:	gap of unknown length
*			contig of 770 bp in length
*	38804	39592:	gap of unknown length
*			contig of 789 bp in length
*	39593	40368:	gap of unknown length
*			contig of 776 bp in length
*	40369	41140:	gap of unknown length
*			contig of 772 bp in length
*	41141	41908:	gap of unknown length
*			contig of 768 bp in length
*	41909	42655:	gap of unknown length
*			contig of 747 bp in length
*	42656	43411:	gap of unknown length
*			contig of 756 bp in length
*	43412	44174:	gap of unknown length
*			contig of 763 bp in length
*	44175	44955:	gap of unknown length
*			contig of 781 bp in length
*	44956	45744:	gap of unknown length
*			contig of 789 bp in length
*	45745	46503:	gap of unknown length
*			contig of 759 bp in length
*	46504	47289:	gap of unknown length
*			contig of 786 bp in length
*	47290	48044:	gap of unknown length
*			contig of 755 bp in length
*	48045	48810:	gap of unknown length
*			contig of 766 bp in length
*	48811	49576:	gap of unknown length
*			contig of 766 bp in length
*	49577	50329:	gap of unknown length
*			contig of 753 bp in length
*	50330	51113:	gap of unknown length
*			contig of 784 bp in length
*	51114	51909:	gap of unknown length
*			contig of 796 bp in length
*			gap of unknown length

* 51910	contig of 841 bp in length
* *	gap of unknown length
* 52751	contig of 907 bp in length
* *	gap of unknown length
* 53658	contig of 790 bp in length
* *	gap of unknown length
* 54448	contig of 766 bp in length
* *	gap of unknown length
* 55214	contig of 777 bp in length
 Query Match 9.18; Score 41.2; DB 55; Length 68523; Best Local Similarity 49.1%; Pred.No.7.5; Matches 114; Conservative 0; Mismatches 117; Indels 1; Gaps 1;	
Qy 213	CCAGGGGACCTGGTACTTGCCCGGAGGCTTCGCCGTGCACGGGTGCACACTGTGGCT 272
Db 53470	CCGCGNGGCCCGCGCCGCGGGGGCGGNGCGGNGCGCGGGCGCGGGCGCGGGCNC 53411
Qy 273	CGCGCTGTGGCTGTGGATGTGGCGCGGAGACACATGTCACTGCCAGTGC-GGGC 331
Db 53410	CGCGCGGGGGCGGGGNGCGGGGGCGCGGGCGCGGGNNCGCGGGGGCGGGGG 53351
Qy 332	ATGGACTGGACCGGCGCGTGTCTGTGTGAGCCCTGAGTCTGC-GCGCACGGCGTG 391
Db 53350	GGGGGGGGGGNNCGCGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGCGG 53291
Qy 392	CACAGCGCGGGGAGGGCGCTCCAGGTCCGAGGGGTTCGGGGGAGCTGG 443
Db 53290	GGGGCGCGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGCGG 53239
 RESULT 15	
LOCUS AC005445	75135 bp DNA HTG 10-FEB-2000
DEFINITION Drosophila melanogaster chromosome 2 clone DS07435 (D359) map	
	44A1-44A2 strain Y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 6
ACCESSION AC005445	unordered pieces.
VERSION AC005445.4	GI:6957993
KEYWORDS HTG; HTGS_PHASE1.	
SOURCE fruit fly.	
ORGANISM Drosophila melanogaster.	
REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
AUTHORS Pterygota; Neuroptera; Endopterygota; Diptera; Brachycera;	
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
(bases 1 to 75135)	
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,	
Butenhof,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,	
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,	
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,	
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,	
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,	
Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,	
Woolley,P., Yang,S., Ye,M., Yu,C. and Rubin,G.M.	
Sequencing of Drosophila melanogaster	
Unpublished	
2 (bases 1 to 75135)	
Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,	
Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,	
Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,	
Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,	
Lomtan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,	
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,	
Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,	
Zieran,L.L. and Kimmel,B.E.	
Direct Submission	
Submitted (14-AUG-1998) Drosophila Genome Center, Lawrence Berkeley	
Laboratory, MS 64-121, Berkeley, CA 94720, USA	
On Feb 11, 2000 this sequence version replaced gi:5656727.	
For further information about this sequence, including its location	
and relationship to other sequences, please visit our sequence	
archive web site (http://www.fruitfly.org/sequence/) or send email	
to bdgp@fruitfly.berkeley.edu. All contigs in this submission sent	
 TITLE JOURNAL	
REFERENCE AUTHORS	
COMMENT	

[illegible]

Search completed: May 26, 2000, 00:15:27
Job time: 3157 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2000, 00:28:51 ; Search time 53.36 Seconds
(without alignments)
118.666 Million cell updates/sec

Title: US-09-099-898-2

Perfect score: 581

Sequence: 1 MKALCLLLPVLGLVSSKT.....CHQCAGMDWTGRCRCRQVP 108

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 168808 seqs, 58629743 residues

Total number of hits satisfying chosen parameters: 168808

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : PIR_63.*

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	86	14.8	934	A34372	complement C6 prec
2	86	14.8	1059	T22545	hypothetical prote
3	82.5	14.2	1574	T13954	MEGF6 protein - ra
4	82	14.1	2321	T78549	notch3 protein - h
5	80.5	13.9	456	1 KX80	protein C (activat
6	78	13.4	1106	T18739	hypothetical prote
7	77	13.3	295	JC5559	lectin-B - Virgini
8	75.5	13.0	387	B49175	Motch A protein -
9	75.5	13.0	2139	A33672	crumbs protein - f
10	75.5	13.0	2531	A46019	Notch-1 protein -
11	75	12.9	124	A21761	high-cysteine chor
12	75	12.9	748	T37097	probable secreted
13	74.5	12.8	2555	A40043	notch protein homo
14	74	12.7	2531	S18188	notch protein homo
15	73.5	12.7	289	T23682	hypothetical prote
16	73.5	12.7	660	F70101	primosomal protein
17	73.5	12.7	1620	T27283	hypothetical prote
18	73	12.6	861	A48825	Notch homolog Motc
19	73	12.6	1049	S19421	ATP-dependent perm
20	73	12.6	1700	S08167	Balbani ring 3 pr
21	73	12.6	2437	S42612	transmembrane prot
22	72	12.4	1111	B44018	laminin B2 chain
23	72	12.4	1193	A44018	laminin B2 chain
24	72	12.4	2318	S45306	notch 3 protein -
25	71.5	12.3	105	S23061	chorion protein -
26	71.5	12.3	519	YRHUR2	dopachrome Delta-1
27	71.5	12.3	1647	T32934	hypothetical prote
28	71.5	12.3	2476	T34022	zonahesin - pig
29	71.5	12.3	2703	A24420	notch protein - fr
30	71	12.2	72	S39418	metallothionein 10

ALIGNMENTS

RESULT 1

A34372

Complement C6 precursor - human

C;Species: Homo sapiens (man)

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C;Accession: A34372; A34235; A32109; A31894; A33072

R;Raefliger, J.A.; Tschopp, J.; Vial, N.; Jenne, D.E.

J. Biol. Chem. 264, 18041-18051, 1989

A;Title: Complete primary structure and functional characterization of the sixth comp

A;Reference number: A34372; MUID:90036879

A;Accession: A34372

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-934 <AE>

A;Cross-references: GB:J05064; NID:gl79703; PIDN:AAA51860.1; PID:gl79704

R;DiScipio, R.G.; Hugli, T.E.

J. Biol. Chem. 264, 16197-16206, 1989

A;Title: The molecular architecture of human complement component C6.

A;Reference number: A34235; MUID:89380223

A;Accession: A34235

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-118,'E',120-934 <DIS>

A;Cross-references: GB:J05024; NID:gl87824; PIDN:AAA59668.1; PID:g307228

R;Chakravarti, D.N.; Chakravarti, B.; Parra, C.A.; Muller-Eberhard, H.J.

Proc. Natl. Acad. Sci. U.S.A. 86, 2799-2803, 1989

A;Title: Structural homology of complement protein C6 with other channel-forming prot

A;Reference number: A32109; MUID:89202413

A;Accession: A32109

A;Molecule type: mRNA

A;Residues: 1-118,'E',120-491 <CH2>

A;Cross-references: GB:J04506; NID:g618465; PIDN:AA59433.1; PID:g618466

R;Chakravarti, D.N.; Muller-Eberhard, H.J.

J. Biol. Chem. 263, 18306-18312, 1988

A;Title: Biochemical characterization of the human complement protein C6. Association

A;Reference number: A31894; MUID:89054009

A;Accession: A31894

A;Molecule type: protein

A;Residues: 22-41 <CHA>

R;Hobart, M.J.; Fernie, B.; DiScipio, R.G.

Biochemistry 32, 6198-6205, 1993

A;Title: Structure of the human C6 gene.

A;Reference number: A53072; MUID:93291175

A;Accession: A53072

A;Status: nucleic acid sequence not shown; not compared with conceptual translation

A;Molecule type: DNA

A;Residues: 'GGGSGG',1-47,'A',49-118,'E',120-561,'A',563-618,'A',620-700,'A',702-763

A;Note: sequence extracted from NCBI backbone (NCBIP:134071)

C;Genetics:

A;Gene: GDB:C6

A;Cross-references: GDB:119045; OMIM:217050

A;Map position: 5p13-5p13

C;Superfamily: complement c6; agrin inhibitor-like repeat homology; complement factor

```

A:Residues: 1-1574 <NAK>
A:Cross-references: EMBL:AB011532; NID:dl226538; PID:dl033425; PIDN:BAA32462.1
A:Experimental source: Strain Sprague-Dawley; brain
C:Genetics:
A:Gene: MEGF6

      Query Match      14.2%; Score 82.5; DB 2; Length 1574;
      Best Local Similarity 35.6%; Pred. No. 4.6;
      Matches 21; Conservative 3; Mismatches 26; Indels 9; Gaps

QY 47 IGLEQSVTSRGDLATCPRGFAVTGCTCGSAGCSWDVRAETTCCHQCACAGMDWTGACCR 105
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 813 VGRQCD-----TCSAGWYGTGCQIRACANDGHCDPTTGRCSCA-PGWIGLSQOR 862

RESULT 4
S78549
notch3 protein - human
C:Species: Homo sapiens (man)
C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 20-Sep-1999
C:Accession: S78549; S71825
R:Joutel, A.; Tournier-Lasserre, E.
submitted to the EMBL Data Library, April 1997
A:Reference number: S78549
A:Accession: S78549
A:Molecule type: mRNA
A:Residues: 1-2321 <JOUL>
A:Cross-references: EMBL:U97669; NID:g2668591; PIDN:AAB91371.1; PID:g2668592
R:Joutel, A.; Corpechot, C.; Ducros, A.; Vahedi, K.G.; Chabriat, H.; Mouton, P.;
x, M.M.; Weissenbach, J.; Bach, J.F.; Bousser, M.G.; Tournier-Lasserre, E.
Nature 383, 707-710, 1996
A:Title: Notch3 mutations
in CADASIL, a hereditary adult-onset condition causi

```

F:133-133/Domain: EGF homology <EGF1>
F:162-194/Domain: EGF homology <EGF1>
F:240-271/Domain: EGF homology <EGX2>
F:318-349/Domain: EGF homology <EGF>
F:473-504/Domain: EGF homology <EGX3>
F:853-884/Domain: EGF homology <EGF3>
F:938-959/Domain: EGF homology <EGX4>
F:1070-1126/Domain: laminin-type EGF-like homology <LEG>
F:1838-1870/Domain: ankyrin repeat homology <AN1>
F:1871-1903/Domain: ankyrin repeat homology <AN2>
F:1905-1937/Domain: ankyrin repeat homology <AN3>
F:1938-1970/Domain: ankyrin repeat homology <AN4>
F:1971-2003/Domain: ankyrin repeat homology <AN5>

```

Query Match      14.1%; Score 82; DB 2; Length 2321;
Best Local Similarity 29.9%; Pred. No. 7;
Matches 23; Conservative 7; Mismatches 23; Indels 24; Gaps
4

Qy 47 IGLEQSVTSRGDLATCPRGFAVGTCTGSGACGSDVRAETTC----- 90
      : : | | | | | : : | | | | : |
Db 1300 VGVPCQQ -TPRGPRCACPQG--LSGPGSCRSPGSPGASNASCAAPCLHGGSCRAPLA 1356
      : : | | | | | : : | | | | : |

Qy 90 ---HCQCAGMDWTGARC 103
      | | | | |
Db 1357 PFPRCACA-QGWTGPRC 1372
      | | | | |

```

RESULT 5

KXBO
protein C (activated) (EC 3.4.21.69) precursor - bovine (fragment)
N:Alternate names: autoprothrombin IIA; plasma protein C
C:Species: Bos primigenius taurus (cattle)
C:Date: 30-Nov-1980 #sequence_revision 17-Mar-1987 #text_change 16-Jul-1999
C:Accession: A26250; A18385; A18386; A00928
R:Long, G.L.; Balagaje, R.M.; MacGillivray, R.T.A.
Proc. Natl. Acad. Sci. U.S.A. 81, 5553-5556, 1984
A:Title: Cloning and sequence of liver cDNA coding for bovine protein C.
A:Reference number: A26250; MUID:85014826
A:Accession: A26250
A:Molecule type: mRNA
A:Residues: 1-456 <LON>
R:Fernlund, P.; Stenflo, J.
J. Biol. Chem. 257, 12170-12179, 1982
A:Title: Amino acid sequence of the light chain of bovine protein C.
A:Reference number: A18385; MUID:83007325
A:Accession: A18385
A:Molecule type: protein
A:Residues: 40-194 <FER>
A:Note: 82-Lys was also found
R:Drakenberg, T.; Fernlund, P.; Roepstorff, P.; Stenflo, J.
Proc. Natl. Acad. Sci. U.S.A. 80, 1802-1806, 1983
A:Title: beta-Hydroxyaspartic acid in vitamin K-dependent protein C.
A:Reference number: A19316; MUID:83169769
A:Contents: annotation; revision to residue 110
R:Stenflo, J.; Fernlund, P.
J. Biol. Chem. 257, 12180-12190, 1982
A:Title: Amino acid sequence of the heavy chain of bovine protein C.
A:Reference number: A18386; MUID:83007326
A:Accession: A18386
A:Molecule type: protein
A:Residues: 197-454, 'PV' <STE>
R:Esmon, N.L.; DeBault, L.E.; Esmon, C.T.
J. Biol. Chem. 258, 5548-5553, 1983
A:Title: Proteolytic formation and properties of gamma-carboxyglutamic acid-domainless P
A:Reference number: A37541; MUID:83213513
R:Johnson, A.E.; Esmon, N.L.; Laue, T.M.; Esmon, C.T.
J. Biol. Chem. 258, 5554-5560, 1983
A:Title: Structural changes required for activation of protein C are induced by Ca2+ bin
A:Reference number: A37542; MUID:83213514
A:Contents: annotation; activation; calcium binding
C:Comment: Protein C is the zymogen of the vitamin K-dependent serine proteinase that re
s.
C:Comment: Protein C is synthesized in the liver as a single chain precursor, which is c
bin, which cleaves a tetradecapeptide from the amino end of the heavy chain; this reacti
C:Comment: Calcium binds to the gamma-carboxyglutamic acid (Gla) residues and, with str
cognition of the thrombin-thrombomodulin complex.
C:Comment: The gamma-carboxyglutamic acid residues arise by a posttranslational, vitamin
C:Superfamily: coagulation factor X; EGF homology; Gla domain homology; trypsin homology
C:Keywords: anticoagulant; beta-hydroxyaspartic acid; blood coagulation; calcium binding
F:1-29/Domain: signal sequence (fragment) #status predicted <SIG>
F:24-83/Domain: Gla domain homology <Gla>
F:30-39/Domain: propeptide #status predicted <PRO>
F:40-194/Product: protein C light chain #status experimental <LCH>
F:98-128/Domain: EGF homology <EG1>
F:137-172/Domain: EGF homology <EG2>
F:197-456/Product: protein C heavy chain #status experimental <HCH>
F:197-210/Domain: activation peptide #status experimental <APT>
F:211-440/Domain: trypsin homology <TRY>
F:45-16, 53, 55, 58, 59, 62, 64, 65, 68, 74/Modified site: gamma-carboxyglutamic acid (Glu) #stat
F:110/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status experimental
F:119-128, 137-148, 144-157, 159-172, 180-318, 237-253, 368-382, 393-421/Disulfide bonds: #stat
F:136, 289, 350/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:252, 298, 397/Active site: His, Asp, Ser #status predicted
F:366/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 13.9%; Score 80.5; DB 1; Length 456;
Best Local Similarity 24.8%; Pred. No. 2.5;

Matches 26; Conservative 10; Mismatches 40; Indels 29; Gaps 4;

QY 18 SKTLCSEEAINEIRIQVAGSLIFRAISSIGLEQSVTSRG--DLATCPRGFAVTG---- 72

Db 57 SEEVCEFEAA-REIFQNTEDTMAFWYSYDGDQCEDRPSGSPCDLPCCGRGKCIDLGGF 115

QY 72 -CTC-----GACGSGWDVRAETTCQCA 94

Db 116 RCDCAEGWGFCHLEVRFSNCRAENGCAHYCMEEGRRHSCA 160

RESULT 6

T18739

hypothetical protein B0393.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T18739

R:Sulston, J.

submitted to the EMBL Data Library, September 1994

A:Reference number: Z19013

A:Accession: T18739

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1106 <WIL>

A:Cross-references: EMBL:Z37983; NID:el519039; PIDN:CAA86058.1; GSPDB:GN00021; CESP:4

A:Experimental source: clone B0393

C:Genetics:

A:Gene: CESP:B0393.5

A:Map position: 3

A:Introns: 15/3; 60/1; 100/3; 343/2; 443/1; 507/3; 609/3; 796/1; 876/1; 945/1;

Query Match 13.4%; Score 78; DB 2; Length 1106;

Best Local Similarity 31.1%; Pred. No. 9;

Matches 19; Conservative 4; Mismatches 22; Indels 16; Gaps 3;

QY 59 DLATCPRG-FAVTGCTGSGAGSGMDVRAETTC--QCAG---MDWTGAR 102

Db 772 DLQTCPRGELGELCDREACRAGHGINCSTCHDCGSVACDVITGCMFALCRAGWESS 831

QY 103 C 103

Db 832 C 832

RESULT 7

JC5559

lectin-B - Virginian pokeweed

C:Species: Phytolacca americana (Virginian pokeweed)

C:Date: 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change 12-Dec-1997

C:Accession: JC5559

R:Yamaguchi, K.; Iurino, N.; Kino, M.; Ishiguro, M.; Funatsu, G.

Biosci. Biotechnol. Biochem. 61, 690-698, 1997

A:Title: The amino acid sequence of mitogenic lectin-B from the roots of pokeweed (Ph

A:Reference number: JC5559; MUID:97290889

A:Accession: JC5559

A:Molecule type: protein

A:Residues: 1-295 <YAM>

A:Experimental source: root

C:Comment: This protein is a lectin specific for N-acetylglucosamine-containing saccha

C:Keywords: glycoprotein

F:96, 139/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 13.3%; Score 77; DB 2; Length 295;

Best Local Similarity 34.6%; Pred. No. 3.7;

Matches 18; Conservative 4; Mismatches 18; Indels 12; Gaps 3;

QY 62 TCRPGFAVTGCTGSGAGSGMDVRAETTCQCC-----AGMDWTG-----ARCC 104

Db 223 TCPNELA---CSSGGWGSNDAHGCGKGCQSCQDYWRGCVDFSGVCPQGRCC 271

QY 16 VSKTLCSMEEA¹NERIQEVAGSLIFRAISSIGLEQCQSVTSRGDLATCPRGFAVTCG 75

Query Match 13.0%; Score 75.5; DB 2; Length 2531;
Best Local Similarity 29.1%; Pred. No. 30;
Matches 2; Conservative 4; Mismatches 24; Indels 33; Gaps 5;
QV 48 GLECSV-----TSRGDLATCPRGFVNTGCTC---GSACGSWDVVR 84

notch protein homolog TAN-1 precursor - human

54 -----VTSRGDL-----ATCPRGFAVTGC--TCGSAC-----GSDVRAETT 88

Db 65 STPEKNAGTCYVDHGGIVDYACSPGLCLTPLANACLANPCRNNGGTCDLLTLE 124
 QY 89 CHCQCAGMDWTGACRCRVQP 108
 Db 125 YKCRCP-PGWSGKSCQQADP 143

RESULT 15

T25682

hypothetical protein F08D12.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T25682

R:Le, T.: Waterston, R.

submitted to the EMBL Data Library, December 1996

A:Description: The sequence of C. elegans cosmid F08D12.

A:Reference number: Z20068

A:Accession: T25682

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-289 <LET>

A:Cross-references: EMBL:U0840; PIDN:AB37924.1; GSPDB:GN00020; CESP:F08D12.3

A:Experimental source: strain Bristol N2; clone F08D12

C:Genetics:

A:Gene: CESP:F08D12.3

A:Map position: 2

A:Introns: 17/1; 56/1; 93/1; 134/1; 171/1; 213/1; 248/1

Query Match

Best Local Similarity 12.7%; Score 73.5; DB 2; Length 289;

Matches 27; Conservative 9; Mismatches 37; Indels 21; Gaps 4;

QY 13 GLLVSKTLCSMEEAINEIRIQEVAGSLIFRAISSIGLEQSVTSRGLATCPRGFAYTGC 72

Db 117 GASITKGSVCCW---NEGIQDAPAS-----ACKSSASPAILGICPSGQVLIGN 161

QY 73 TCGSACGSWDYRAETTCQCQ---AGMDWTGARC 103

Db 162 YC---CEAKDVPDPSTATCQGVYAGPYGTGLTC 192

Search completed: May 26, 2000, 03:18:28

Job time: 10177 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2000, 03:01:52 ; Search time 40.45 Seconds
(without alignments)

81.314 Million cell updates/sec

Title: US-09-099-898-2

Perfect score: 581

Sequence: 1 MKALCLLLPVGLVSSKT.....CHCQAGMDWGRCRCVQP 108

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 83857 seqs, 30454973 residues

Total number of hits satisfying chosen parameters: 83857

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt_38.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	86	14.8	934	1 CO6_HUMAN	P13671 homo sapien
2	80.5	13.9	456	1 PRTC_BOVIN	P00745 bos taurus
3	75.5	13.0	2139	1 CRB_DROME	P10040 drosophila
4	75.5	13.0	2531	1 NTC1_MOUSE	Q01705 mus musculus
5	74.5	12.8	2444	1 NTC1_HUMAN	P46531 homo sapien
6	74	12.7	124	1 CHH2_BOMMO	P05687 bombyx mori
7	74	12.7	2531	1 NTC1_RAT	Q07008 rattus norv
8	73.5	12.7	660	1 PRIA_BORBU	Q45032 borrelia bu
9	73	12.6	1049	1 ADP1_YEAST	P25371 saccharomyc
10	73	12.6	1700	1 BAR3_CHITE	Q03376 chironomus
11	73	12.6	2437	1 NOTC_BRARE	P46530 brachydanio
12	72	12.4	1193	1 LMG2_HUMAN	Q13753 homo sapien
13	72	12.4	2318	1 NTC3_MOUSE	Q61982 mus musculus
14	71.5	12.3	519	1 TYR2_HUMAN	P40126 homo sapien
15	71.5	12.3	2476	1 ZAN_PIG	Q28983 sus scrofa
16	71.5	12.3	2703	1 NOTC_DROME	P07207 drosophila
17	71	12.2	72	1 MT13_MYED	P80248 mytilus edu
18	71	12.2	397	1 PRLA_LIXEN	P00778 lysobacter
19	71	12.2	1134	1 TIE1_MOUSE	Q06806 mus musculus
20	71	12.2	2907	1 FBN2_MOUSE	Q61555 mus musculus
21	70	12.0	2871	1 FBN1_MOUSE	Q61554 mus musculus
22	69.5	12.0	2524	1 NOTC_XENLA	P21783 xenopus lae
23	69	11.9	2715	1 G156_PAPPR	P13837 paramecium
24	68.5	11.8	581	1 IRR_RAT	Q64716 rattus norv
25	68	11.7	72	1 MT11_MYED	P80246 mytilus edu
26	68	11.7	458	1 PRTC_RABIT	Q28661 oryctolagus
27	68	11.7	1138	1 TIE1_HUMAN	P35590 homo sapien
28	67	11.5	72	1 MT12_MYED	P80247 mytilus edu
29	67	11.5	364	1 PGLR_COCCA	P26215 cochllobolu
30	67	11.5	372	1 AGI_URTDI	P11218 urtica dioi
31	67	11.5	1136	1 TIE1_BOVIN	Q06805 bos taurus
32	67	11.5	1801	1 LMB2_RAT	P15800 rattus norv
33	66.5	11.4	71	1 MT21_MYED	P80251 mytilus edu
34	66.5	11.4	182	1 HPRT_RHOCA	P37171 rhodobacter

RESULT 1

ID	CO6_HUMAN	STANDARD;	PRT;	934 AA.
AC	P13671;			
DT	01-JAN-1990 (Rel. 13, Created)			
DT	01-AUG-1991 (Rel. 19, Last sequence update)			
DT	15-FEB-2000 (Rel. 39, Last annotation update)			
DE	COMPLEMENT COMPONENT C6 PRECURSOR.			
GN	C6.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Primates; Catarrhini; Hominidae; Homo.			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 22-31 AND 633-640.			
RX	MEDLINE; 90036879.			
RA	Haeffliger J.-A., Tschopp J., Vial N., Jenne D.E.;			
RT	"Complete primary structure and functional characterization of the			
RT	sixth component of the human complement system. Identification of the			
RT	C5b-binding domain in complement C6.";			
RL	J. Biol. Chem. 264:18041-18051(1989).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 89380223.			
RA	Discipio R.G., Hugli T.E.;			
RT	"The molecular architecture of human complement component C6.";			
RL	J. Biol. Chem. 264:16197-16206(1989).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-BLOOD.			
RX	MEDLINE; 93291175.			
RA	Hobart M.J., Fernie B., Discipio R.G.;			
RT	"Structure of the human C6 gene.";			
RL	Biochemistry 32:6198-6205(1993).			
RN	[4]			
RP	SEQUENCE OF 1-491 FROM N.A.			
RX	MEDLINE; 89202413.			
RA	Chakravarti D.N., Chakravarti B., Parra C.A., Mueller-Eberhard H.J.;			
RT	"Structural homology of complement protein C6 with other			
RT	channel-forming proteins of complement.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 86:2799-2803(1989).			
RN	[5]			
RP	VARIANT ALLOTYPES C6 A AND B.			
RX	MEDLINE; 93326158.			
RA	Dewald G., Nothen M.M., Cichon S.;			
RT	"Polymorphism of human complement component C6: an amino acid			
RT	substitution (Glu/Ala) within the second thombospondin repeat			
RT	differentiates between the two common allotypes C6 A and C6 B.";			
RL	Biochem. Biophys. Res. Commun. 194:458-464(1993).			
RN	[6]			
RP	CARBOHYDRATE-BINDING SITES.			
RX	MEDLINE; 20020247.			
RA	Hofsteenge J., Blommers M., Hess D., Furmanek A., Miroshnichenko O.;			
RT	"The four terminal components of the complement system are			
RT	C-mannosylated on multiple tryptophan residues.";			
RL	J. Biol. Chem. 274:32786-32794(1999).			
CC	-1- FUNCTION: INVOLVED IN THE FORMATION OF THE LYtic C5B-9M COMPLEX.			
CC	-1- DOMAIN: C6 CONTAINS NINE DISTINCT STRUCTURAL UNITS (CYSTEINE-RICH)			

35	66.5	11.4	444	1 SHU4_ECOLI	P03748 escherichia
36	66.5	11.4	1416	1 YN81_CAEEL	Q03610 caenorhabdi
37	66.5	11.4	2911	1 FBN2_HUMAN	P35556 homo sapien
38	66	11.4	72	1 MT14_MYED	P80249 mytilus edu
39	66	11.4	177	1 B7C_MOUSE	Q05928 mus musculus
40	66	11.4	309	1 PNAD_PIG	Q28955 sus scrofa
41	66	11.4	463	1 MFGM_MOUSE	P21956 mus musculus
42	66	11.4	1300	1 IRR_CAVPO	P14617 cavia porce
43	66	11.4	2871	1 FBN1_HUMAN	P35555 homo sapien
44	65.5	11.3	66	1 MT_ARIAR	P55946 arianta arb
45	65.5	11.3	71	1 MT22_MYED	P80252 mytilus edu

ALIGNMENTS

RT carboxylglutamic acid";
 RL J. Biol. Chem. 258:5554-5560(1983).
 CC -!- FUNCTION: PROTEIN C IS A VITAMIN K-DEPENDENT SERINE PROTEASE THAT
 CC REGULATES BLOOD COAGULATION BY INACTIVATING FACTORS VA AND VIIIA
 CC IN THE PRESENCE OF CALCIUM IONS AND PHOSPHOLIPIDS.
 CC -!- CATALYTIC ACTIVITY: DEGRADATION OF BLOOD COAGULATION FACTORS VA
 CC AND VIIIA.
 CC -!- SUBUNIT: SYNTHESIZED AS A SINGLE CHAIN PRECURSOR, WHICH IS CLEAVED
 CC INTO A LIGHT CHAIN AND A HEAVY CHAIN HELD TOGETHER BY A DISULFIDE
 CC BOND. THE ENZYME IS THEN ACTIVATED BY THROMBIN, WHICH CLEAVES A
 CC TETRAPEPTIDE FROM THE AMINO END OF THE HEAVY CHAIN; THIS
 CC REACTION, WHICH OCCURS AT THE SURFACE OF ENDOTHELIAL CELLS, IS
 CC STRONGLY PROMOTED BY THROMBOMODULIN.
 CC -!- TISSUE SPECIFICITY: PLASMA; SYNTHESIZED IN THE LIVER.
 CC -!- PTM: THE VITAMIN K-DEPENDENT, ENZYMAIC CARBOXYLATION OF SOME
 CC GLU RESIDUES ALLOWS THE MODIFIED PROTEIN TO BIND CALCIUM.
 CC -!- MISCELLANEOUS: CALCIUM ALSO BINDS, WITH STRONGER AFFINITY TO
 CC ANOTHER SITE, BEYOND THE GLA DOMAIN. THIS GLA-INDEPENDENT BINDING
 CC SITE IS NECESSARY FOR THE RECOGNITION OF THE
 CC THROMBIN-THROMBOMODULIN COMPLEX.
 CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: K02435; AAA30685.1; -
 DR PIR: A00928; KXBO
 DR HSP: P04070; LAUT.
 DR PROSITE: PS00010; ASX_HYDROXYL; 1.
 DR PROSITE: PS00011; GLU CARBOXYLATION; 1.
 DR PROSITE: PS00022; EGF_1; 1.
 DR PROSITE: PS01186; EGF_2; 2.
 DR PROSITE: PS01187; EGF_CA; 1.
 DR PROSITE: PS00134; TRYPsin_HIS; FALSE_NEG.
 DR PROSITE: PS00135; TRYPsin_SER; 1.
 DR PFAM: PF00008; EGF; 2.
 DR PFAM: PF00089; trypsin; 1.
 DR PFAM: PF00594; gla; 1.
 DR KW Blood coagulation; Glycoprotein; Plasma; Serine protease; Liver;
 KW Gamma-carboxylglutamic acid; Calcium-binding; Vitamin K; Hydroxylation;
 KW EGF-like domain; Repeat; Endothelial cell; Hydrolase; Signal.
 FT NON_TER 1 1
 FT SIGNAL <1 29
 FT PROPEP 30 29
 FT CHAIN 40 194
 FT CHAIN 197 456
 FT PEPTIDE 197 210
 FT DOMAIN 94 129
 FT DOMAIN 133 173
 FT DOMAIN 211 456
 FT MOD_RES 45 45
 FT MOD_RES 46 46
 FT MOD_RES 53 53
 FT MOD_RES 55 55
 FT MOD_RES 58 58
 FT MOD_RES 59 59
 FT MOD_RES 62 62
 FT MOD_RES 64 64
 FT MOD_RES 65 65
 FT MOD_RES 68 68
 FT MOD_RES 74 74
 FT MOD_RES 110 110
 FT ACT_SITE 252 252
 FT ACT_SITE 298 298
 FT ACT_SITE 397 397
 FT DISULFID 56 61

FT DISULFID 89 108
 FT DISULFID 98 103
 FT DISULFID 102 117
 FT DISULFID 119 128
 FT DISULFID 137 148
 FT DISULFID 144 157
 FT DISULFID 159 172
 FT DISULFID 180 318
 FT DISULFID 237 253
 FT DISULFID 368 382
 FT DISULFID 393 421
 FT CARBOHYD 136 136
 FT CARBOHYD 289 289
 FT CARBOHYD 350 350
 FT CARBOHYD 366 366
 FT VARIANT 82 82
 FT CONFLICT 455 456
 SQ SEQUENCE 456 AA; 51407 MW; CAAF6833F894C209 CRC64;
 ASN-X-CYS (INSTEAD OF SER/THR).
 F -> K.
 VP -> PV (IN REF. 4).
 Query Match 13.9%; Score 80.5; DB 1; Length 456;
 Best Local Similarity 24.8%; Pred. No. 0.61;
 Matches 26; Conservative 10; Mismatches 40; Indels 29; Gaps 4;
 QY 18 SKTLCSEEAINEIRIQEVAGSLIPRAISSIGLEQSVTSRG--DLATPRGFAVIG---- 72
 Db 57 SEEVCEFEAA-REIFQNTDMTFWFSYDGDQDRPSGCDLPCCGRKCIDGLGFG 115
 QY 72 -CTC-----GSACGSWDVRAETTCQCQA 94
 Db 116 RDCAEGWGRFCLHEVRFSNCSAENGCAHYCEEGRRHCSCA 160
 RESULT 3
 CRB_DROME
 ID CRB_DROME STANDARD; PRY: 2139 AA.
 AC P10040;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE CRUMBS PROTEIN PRECURSOR (95F).
 GN CRB.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OREGON-R; TISSUE-EMBRYO;
 RX MEDLINE; 90263104.
 RA Tepass U., Theres C., Knust E.;
 RT "Crums encodes an EGF-like protein expressed on apical membranes of
 RT Drosophila epithelial cells and required for organization of
 RT epithelia";
 RL Cell 61:787-799(1990).
 RN [2]
 RP SEQUENCE OF 1663-1955 FROM N.A.
 RX MEDLINE; 87218537.
 RA Knust E., Dietrich U., Tepass U., Bremer K.A., Weigel D.,
 RA Vaessin H., Campos-Ortega J.A.;
 RT "EGF homologous sequences encoded in the genome of Drosophila
 RT melanogaster, and their relation to neurogenic genes";
 RL EMBO J. 6:761-766(1987).
 CC -!- FUNCTION: MAY PLAY A ROLE IN THE DEVELOPMENT OF EPITHELIA,
 CC POSSIBLY FOR THE ESTABLISHMENT AND/OR MAINTENANCE OF CELL
 CC POLARITY. IT MAY ACT AS A SIGNAL.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- PTM: PHOSPHORYLATED IN THE CYTOPLASMIC DOMAIN (POTENTIAL).
 CC -!- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M33753; AAA28428.1; ALT_SEQ.
DR EMBL; X05144; CAA28793.1; -
DR PIR; B26637; B26637.
DR PIR; A35672; A35672.
DR HSP; P00740; 11XA.
DR FLYBASE; FBgn0000368; crb.
DR PRINTS; PR00010; EGFBLD.
DR PROSITE; PS00010; ASX_HYDROXYL; 15.
DR PROSITE; PS00022; EGF_1; 26.
DR PROSITE; PS01186; EGF_2; 17.
DR PROSITE; PS01187; EGF_CA; 15.
DR PFAM; PF00008; EGF; 27.
DR PFAM; PF00054; laminin_G; 3.
KW Differentiation; Repeat; EGF-like domain; Transmembrane;
KW Glycoprotein; Signal; Phosphorylation.
FT SIGNAL 1 90
FT CHAIN 91 2139 CRUMB PROTEIN.
FT DOMAIN 91 2084 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 2085 2111 POTENTIAL.
FT DOMAIN 2112 2139 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 267 303 EGF-LIKE 1.
FT DOMAIN 306 343 EGF-LIKE 2.
FT DOMAIN 348 386 EGF-LIKE 3.
FT DOMAIN 388 425 EGF-LIKE 4.
FT DOMAIN 427 463 EGF-LIKE 5.
FT DOMAIN 464 500 EGF-LIKE 6.
FT DOMAIN 501 532 EGF-LIKE 7.
FT DOMAIN 545 581 EGF-LIKE 8.
FT DOMAIN 582 611 EGF-LIKE 9.
FT DOMAIN 609 646 EGF-LIKE 10.
FT DOMAIN 648 685 EGF-LIKE 11.
FT DOMAIN 687 723 EGF-LIKE 12.
FT DOMAIN 725 761 EGF-LIKE 13.
FT DOMAIN 763 800 EGF-LIKE 14.
FT DOMAIN 802 838 EGF-LIKE 15.
FT DOMAIN 840 902 EGF-LIKE 16.
FT DOMAIN 904 940 EGF-LIKE 17.
FT DOMAIN 942 978 EGF-LIKE 18.
FT DOMAIN 980 1021 EGF-LIKE 19.
FT DOMAIN 1207 1243 EGF-LIKE 20.
FT DOMAIN 1481 1517 EGF-LIKE 21.
FT DOMAIN 1759 1795 EGF-LIKE 22.
FT DOMAIN 1797 1833 EGF-LIKE 23.
FT DOMAIN 1835 1871 EGF-LIKE 24.
FT DOMAIN 1874 1915 EGF-LIKE 25.
FT DOMAIN 1915 1951 EGF-LIKE 26.
FT DOMAIN 1953 1989 EGF-LIKE 27.
FT DOMAIN 1991 2029 EGF-LIKE 28.
FT DOMAIN 2030 2070 EGF-LIKE 29.
FT DISULFID 271 282 BY SIMILARITY.
FT DISULFID 276 291 BY SIMILARITY.
FT DISULFID 293 302 BY SIMILARITY.
FT DISULFID 310 321 BY SIMILARITY.
FT DISULFID 315 331 BY SIMILARITY.
FT DISULFID 333 342 BY SIMILARITY.
FT DISULFID 352 363 BY SIMILARITY.
FT DISULFID 357 374 BY SIMILARITY.
FT DISULFID 376 385 BY SIMILARITY.
FT DISULFID 392 403 BY SIMILARITY.
FT DISULFID 397 412 BY SIMILARITY.
FT DISULFID 414 424 BY SIMILARITY.
FT DISULFID 431 442 BY SIMILARITY.
FT DISULFID 436 451 BY SIMILARITY.
FT DISULFID 453 462 BY SIMILARITY.
FT DISULFID 468 479 BY SIMILARITY.
FT DISULFID 473 488 BY SIMILARITY.
FT DISULFID 490 499 BY SIMILARITY.
FT DISULFID 505 515 BY SIMILARITY.
FT DISULFID 509 520
FT DISULFID 522 531 BY SIMILARITY.
FT DISULFID 549 562 BY SIMILARITY.
FT DISULFID 556 569 BY SIMILARITY.
FT DISULFID 571 580 BY SIMILARITY.
FT DISULFID 586 597 BY SIMILARITY.
FT DISULFID 591 602 BY SIMILARITY.
FT DISULFID 604 610 BY SIMILARITY.
FT DISULFID 613 624 BY SIMILARITY.
FT DISULFID 618 634 BY SIMILARITY.
FT DISULFID 636 645 BY SIMILARITY.
FT DISULFID 652 664 BY SIMILARITY.
FT DISULFID 659 673 BY SIMILARITY.
FT DISULFID 675 684 BY SIMILARITY.
FT DISULFID 691 702 BY SIMILARITY.
FT DISULFID 696 711 BY SIMILARITY.
FT DISULFID 713 722 BY SIMILARITY.
FT DISULFID 729 740 BY SIMILARITY.
FT DISULFID 734 749 BY SIMILARITY.
FT DISULFID 751 760 BY SIMILARITY.
FT DISULFID 767 778 BY SIMILARITY.
FT DISULFID 772 787 BY SIMILARITY.
FT DISULFID 789 799 BY SIMILARITY.
FT DISULFID 806 817 BY SIMILARITY.
FT DISULFID 811 826 BY SIMILARITY.
FT DISULFID 828 837 BY SIMILARITY.
FT DISULFID 844 855 BY SIMILARITY.
FT DISULFID 849 890 BY SIMILARITY.
FT DISULFID 892 901 BY SIMILARITY.
FT DISULFID 908 919 BY SIMILARITY.
FT DISULFID 913 928 BY SIMILARITY.
FT DISULFID 930 939 BY SIMILARITY.
FT DISULFID 946 957 BY SIMILARITY.
FT DISULFID 952 966 BY SIMILARITY.
FT DISULFID 968 977 BY SIMILARITY.
FT DISULFID 984 995 BY SIMILARITY.
FT DISULFID 989 1009 BY SIMILARITY.
FT DISULFID 1011 1020 BY SIMILARITY.
FT DISULFID 1211 1222 BY SIMILARITY.
FT DISULFID 1216 1231 BY SIMILARITY.
FT DISULFID 1233 1242 BY SIMILARITY.
FT DISULFID 1485 1496 BY SIMILARITY.
FT DISULFID 1490 1505 BY SIMILARITY.
FT DISULFID 1507 1516 BY SIMILARITY.
FT DISULFID 1763 1774 BY SIMILARITY.
FT DISULFID 1768 1783 BY SIMILARITY.
FT DISULFID 1785 1794 BY SIMILARITY.
FT DISULFID 1801 1812 BY SIMILARITY.
FT DISULFID 1806 1821 BY SIMILARITY.
FT DISULFID 1823 1832 BY SIMILARITY.
FT DISULFID 1839 1850 BY SIMILARITY.
FT DISULFID 1844 1859 BY SIMILARITY.
FT DISULFID 1861 1870 BY SIMILARITY.
FT DISULFID 1878 1889 BY SIMILARITY.
FT DISULFID 1883 1903 BY SIMILARITY.
FT DISULFID 1905 1914 BY SIMILARITY.
FT DISULFID 1919 1930 BY SIMILARITY.
FT DISULFID 1924 1939 BY SIMILARITY.
FT DISULFID 1941 1950 BY SIMILARITY.
FT DISULFID 1957 1968 BY SIMILARITY.
FT DISULFID 1962 1977 BY SIMILARITY.
FT DISULFID 1979 1988 BY SIMILARITY.
FT DISULFID 1995 2008 BY SIMILARITY.
FT DISULFID 2002 2017 BY SIMILARITY.
FT DISULFID 2019 2028 BY SIMILARITY.
FT CARBOHYD 37 36 POTENTIAL.
FT CARBOHYD 96 96 POTENTIAL.
FT CARBOHYD 198 198 POTENTIAL.
FT CARBOHYD 238 238 POTENTIAL.
FT CARBOHYD 239 239 POTENTIAL.
FT CARBOHYD 336 336 POTENTIAL.
FT CARBOHYD 400 400 POTENTIAL.
FT CARBOHYD 550 550 POTENTIAL.


```
FT DISULFID 68 BY SIMILARITY.
FT DISULFID 89 BY SIMILARITY.
FT DISULFID 106 BY SIMILARITY.
FT DISULFID 111 BY SIMILARITY.
FT DISULFID 127 BY SIMILARITY.
FT DISULFID 129 BY SIMILARITY.
FT DISULFID 138 BY SIMILARITY.
FT DISULFID 144 BY SIMILARITY.
FT DISULFID 149 BY SIMILARITY.
FT DISULFID 166 BY SIMILARITY.
FT DISULFID 175 BY SIMILARITY.
FT DISULFID 182 BY SIMILARITY.
FT DISULFID 189 BY SIMILARITY.
FT DISULFID 204 BY SIMILARITY.
FT DISULFID 206 BY SIMILARITY.
FT DISULFID 222 BY SIMILARITY.
FT DISULFID 227 BY SIMILARITY.
FT DISULFID 245 BY SIMILARITY.
FT DISULFID 261 BY SIMILARITY.
FT DISULFID 266 BY SIMILARITY.
FT DISULFID 283 BY SIMILARITY.
FT DISULFID 299 BY SIMILARITY.
FT DISULFID 306 BY SIMILARITY.
FT DISULFID 323 BY SIMILARITY.
FT DISULFID 339 BY SIMILARITY.
FT DISULFID 344 BY SIMILARITY.
FT DISULFID 361 BY SIMILARITY.
FT DISULFID 376 BY SIMILARITY.
FT DISULFID 387 BY SIMILARITY.
FT DISULFID 400 BY SIMILARITY.
FT DISULFID 416 BY SIMILARITY.
FT DISULFID 423 BY SIMILARITY.
FT DISULFID 440 BY SIMILARITY.
FT DISULFID 456 BY SIMILARITY.
FT DISULFID 461 BY SIMILARITY.
FT DISULFID 478 BY SIMILARITY.
FT DISULFID 494 BY SIMILARITY.
FT DISULFID 499 BY SIMILARITY.
FT DISULFID 516 BY SIMILARITY.
FT DISULFID 532 BY SIMILARITY.
FT DISULFID 537 BY SIMILARITY.
FT DISULFID 554 BY SIMILARITY.
FT DISULFID 570 BY SIMILARITY.
FT DISULFID 575 BY SIMILARITY.
FT DISULFID 591 BY SIMILARITY.
FT DISULFID 607 BY SIMILARITY.
FT DISULFID 612 BY SIMILARITY.
FT DISULFID 629 BY SIMILARITY.
FT DISULFID 645 BY SIMILARITY.
FT DISULFID 650 BY SIMILARITY.
FT DISULFID 666 BY SIMILARITY.
FT DISULFID 682 BY SIMILARITY.
FT DISULFID 687 BY SIMILARITY.
FT DISULFID 704 BY SIMILARITY.
FT DISULFID 720 BY SIMILARITY.
FT DISULFID 725 BY SIMILARITY.
FT DISULFID 741 BY SIMILARITY.
FT DISULFID 757 BY SIMILARITY.
FT DISULFID 762 BY SIMILARITY.
FT DISULFID 779 BY SIMILARITY.
FT DISULFID 795 BY SIMILARITY.
FT DISULFID 800 BY SIMILARITY.
FT DISULFID 817 BY SIMILARITY.
FT DISULFID 833 BY SIMILARITY.
FT DISULFID 838 BY SIMILARITY.
FT DISULFID 857 BY SIMILARITY.
FT DISULFID 873 BY SIMILARITY.
FT DISULFID 878 BY SIMILARITY.
FT DISULFID 895 BY SIMILARITY.
FT DISULFID 911 BY SIMILARITY.
FT DISULFID 916 BY SIMILARITY.
FT DISULFID 933 BY SIMILARITY.
FT DISULFID 937 BY SIMILARITY.
FT DISULFID 987 BY SIMILARITY.
FT DISULFID 992 BY SIMILARITY.
FT DISULFID 1009 BY SIMILARITY.
FT DISULFID 1025 BY SIMILARITY.
FT DISULFID 1030 BY SIMILARITY.
87 BY SIMILARITY.
98 BY SIMILARITY.
117 BY SIMILARITY.
127 BY SIMILARITY.
138 BY SIMILARITY.
155 BY SIMILARITY.
164 BY SIMILARITY.
175 BY SIMILARITY.
182 BY SIMILARITY.
204 BY SIMILARITY.
225 BY SIMILARITY.
233 BY SIMILARITY.
243 BY SIMILARITY.
254 BY SIMILARITY.
272 BY SIMILARITY.
281 BY SIMILARITY.
292 BY SIMILARITY.
312 BY SIMILARITY.
321 BY SIMILARITY.
332 BY SIMILARITY.
350 BY SIMILARITY.
359 BY SIMILARITY.
370 BY SIMILARITY.
387 BY SIMILARITY.
398 BY SIMILARITY.
409 BY SIMILARITY.
429 BY SIMILARITY.
438 BY SIMILARITY.
449 BY SIMILARITY.
467 BY SIMILARITY.
476 BY SIMILARITY.
487 BY SIMILARITY.
505 BY SIMILARITY.
514 BY SIMILARITY.
525 BY SIMILARITY.
543 BY SIMILARITY.
552 BY SIMILARITY.
563 BY SIMILARITY.
580 BY SIMILARITY.
589 BY SIMILARITY.
600 BY SIMILARITY.
618 BY SIMILARITY.
627 BY SIMILARITY.
638 BY SIMILARITY.
655 BY SIMILARITY.
664 BY SIMILARITY.
675 BY SIMILARITY.
693 BY SIMILARITY.
702 BY SIMILARITY.
713 BY SIMILARITY.
730 BY SIMILARITY.
739 BY SIMILARITY.
750 BY SIMILARITY.
768 BY SIMILARITY.
777 BY SIMILARITY.
788 BY SIMILARITY.
806 BY SIMILARITY.
815 BY SIMILARITY.
826 BY SIMILARITY.
844 BY SIMILARITY.
855 BY SIMILARITY.
866 BY SIMILARITY.
884 BY SIMILARITY.
893 BY SIMILARITY.
904 BY SIMILARITY.
922 BY SIMILARITY.
931 BY SIMILARITY.
942 BY SIMILARITY.
958 BY SIMILARITY.
988 BY SIMILARITY.
1007 BY SIMILARITY.
1018 BY SIMILARITY.
1036 BY SIMILARITY.
1045 BY SIMILARITY.
FT DISULFID 1047 1056 BY SIMILARITY.
FT DISULFID 1063 1074 BY SIMILARITY.
FT DISULFID 1068 1083 BY SIMILARITY.
FT DISULFID 1085 1094 BY SIMILARITY.
FT DISULFID 1101 1122 BY SIMILARITY.
FT DISULFID 1116 1131 BY SIMILARITY.
FT DISULFID 1133 1142 BY SIMILARITY.
FT DISULFID 1149 1160 BY SIMILARITY.
FT DISULFID 1154 1169 BY SIMILARITY.
FT DISULFID 1171 1180 BY SIMILARITY.
FT DISULFID 1187 1198 BY SIMILARITY.
FT DISULFID 1192 1207 BY SIMILARITY.
FT DISULFID 1209 1218 BY SIMILARITY.
Query Match 13.0%; Score 75.5; DB 1; Length 2531;
Best Local Similarity 29.1%; Pred. No. 9;
Matches 25; Conservative 4; Mismatches 24; Indels 33; Gaps 5;
QY 48 GLEQSV-----TSRGDLATCPRGFAVTCGTC---GSACGSWDVR 84
Db 1301 GRRCESVINGCRPKNGVCAVASTARGFCRCPAGE--EGATCENDARTCGSLRCL 1358
QY 85 AETTC-----HCQACAGMDWTGARC 103
Db 1359 NGGTCISGPRSPPTCLCLG-SFTGPEC 1383
RESULT 5
NTCL_HUMAN STANDARD; PRT; 2444 AA.
ID P46531;
DT 01-NOV-1995 (Rel. 32, Created);
DT 01-NOV-1995 (Rel. 32, Last sequence update);
DT 01-FEB-1996 (Rel. 33, Last annotation update);
DE NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1 PRECURSOR (TRANSLATION-
GN NOTCH1 OR TAN1).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP MEDLINE; 91347367.
RX Ellisen L.W., Bird J., West D.C., Soreng A.L., Reynolds T.C.,
RA Smith S.D., Sklar J.;
RT "TAN-1, the human homolog of the Drosophila notch gene, is broken by
chromosomal translocations in T lymphoblastic neoplasms.";
RL Cell 66:649-661(1991).
CC -!- FUNCTION: MAY BE IMPORTANT FOR NORMAL LYMPHOCYTE FUNCTION. IN
ALTERED FORM, MAY CONTRIBUTE TO TRANSFORMATION OR PROGRESSION
IN SOME T-CELL NEOPLASMS.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: IN FETAL TISSUES MOST ABUNDANT IN SPLEEN,
BRAIN STEM AND LUNG. ALSO PRESENT IN MOST ADULT TISSUES WHERE IT
IS FOUND MAINLY IN LYMPHOID TISSUES.
CC -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M73980; AAA60614.1; -.
DR HSP; P00740; IYXA.
DR MIM; 190198; -.
DR PROSITE; PS00010; ASX_HYDROXYL; 20.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS00022; EGF_1; 34.
```


[illegible]


```

FT DOMAIN 218 255 EGF-LIKE 6.
FT DOMAIN 257 293 EGF-LIKE 7. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 295 333 EGF-LIKE 8. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 335 371 EGF-LIKE 9. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 372 410 EGF-LIKE 10.
FT DOMAIN 412 450 EGF-LIKE 11. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 452 488 EGF-LIKE 12. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 490 526 EGF-LIKE 13. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 528 564 EGF-LIKE 14. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 566 601 EGF-LIKE 15. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 603 639 EGF-LIKE 16. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 641 676 EGF-LIKE 17. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 678 714 EGF-LIKE 18. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 716 751 EGF-LIKE 19. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 753 789 EGF-LIKE 20. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 791 827 EGF-LIKE 21. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 829 867 EGF-LIKE 22.
FT DOMAIN 869 905 EGF-LIKE 23. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 907 943 EGF-LIKE 24.
FT DOMAIN 945 981 EGF-LIKE 25. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 983 1019 EGF-LIKE 26. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1021 1057 EGF-LIKE 27. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1059 1095 EGF-LIKE 28.
FT DOMAIN 1097 1133 EGF-LIKE 29.
FT DOMAIN 1145 1181 EGF-LIKE 30. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1183 1219 EGF-LIKE 31. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1221 1265 EGF-LIKE 32. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1267 1305 EGF-LIKE 33.
FT DOMAIN 1307 1346 EGF-LIKE 34.
FT DOMAIN 1348 1384 EGF-LIKE 35.
FT DOMAIN 1387 1426 EGF-LIKE 36.
FT DOMAIN 1449 1462 CYS-RICH.
FT DOMAIN 1465 1499 6 X ANK MOTIF REPEATS.
FT REPEAT 1865 1910 ANK MOTIF 1.
FT REPEAT 1912 1942 ANK MOTIF 2.
FT REPEAT 1944 1975 ANK MOTIF 3.
FT REPEAT 1978 2009 ANK MOTIF 4.
FT REPEAT 2011 2042 ANK MOTIF 5.
FT REPEAT 2044 2076 ANK MOTIF 6.
FT DISULFID 24 37 BY SIMILARITY.
FT DISULFID 31 46 BY SIMILARITY.
FT DISULFID 48 57 BY SIMILARITY.
FT DISULFID 63 74 BY SIMILARITY.
FT DISULFID 68 87 BY SIMILARITY.
FT DISULFID 89 98 BY SIMILARITY.
FT DISULFID 106 117 BY SIMILARITY.
FT DISULFID 111 127 BY SIMILARITY.
FT DISULFID 129 138 BY SIMILARITY.
FT DISULFID 144 155 BY SIMILARITY.
FT DISULFID 149 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 182 195 BY SIMILARITY.
FT DISULFID 189 204 BY SIMILARITY.
FT DISULFID 206 215 BY SIMILARITY.
FT DISULFID 222 233 BY SIMILARITY.
FT DISULFID 227 243 BY SIMILARITY.
FT DISULFID 245 254 BY SIMILARITY.
FT DISULFID 261 272 BY SIMILARITY.
FT DISULFID 266 281 BY SIMILARITY.
FT DISULFID 283 292 BY SIMILARITY.
FT DISULFID 299 312 BY SIMILARITY.
FT DISULFID 306 321 BY SIMILARITY.
FT DISULFID 323 332 BY SIMILARITY.
FT DISULFID 339 350 BY SIMILARITY.
FT DISULFID 344 359 BY SIMILARITY.
FT DISULFID 361 370 BY SIMILARITY.
FT DISULFID 376 387 BY SIMILARITY.
FT DISULFID 381 398 BY SIMILARITY.
FT DISULFID 400 409 BY SIMILARITY.
FT DISULFID 416 429 BY SIMILARITY.
FT DISULFID 423 438 BY SIMILARITY.
FT DISULFID 440 449 BY SIMILARITY.
FT DISULFID 446 467 BY SIMILARITY.

```

```

FT DISULFID 461 476 BY SIMILARITY.
FT DISULFID 478 487 BY SIMILARITY.
FT DISULFID 494 505 BY SIMILARITY.
FT DISULFID 499 514 BY SIMILARITY.
FT DISULFID 516 525 BY SIMILARITY.
FT DISULFID 532 543 BY SIMILARITY.
FT DISULFID 537 552 BY SIMILARITY.
FT DISULFID 554 563 BY SIMILARITY.
FT DISULFID 570 580 BY SIMILARITY.
FT DISULFID 575 589 BY SIMILARITY.
FT DISULFID 591 600 BY SIMILARITY.
FT DISULFID 607 618 BY SIMILARITY.
FT DISULFID 612 627 BY SIMILARITY.
FT DISULFID 629 638 BY SIMILARITY.
FT DISULFID 645 655 BY SIMILARITY.
FT DISULFID 650 664 BY SIMILARITY.
FT DISULFID 666 675 BY SIMILARITY.
FT DISULFID 682 693 BY SIMILARITY.
FT DISULFID 687 702 BY SIMILARITY.
FT DISULFID 704 713 BY SIMILARITY.
FT DISULFID 720 730 BY SIMILARITY.
FT DISULFID 725 739 BY SIMILARITY.
FT DISULFID 741 750 BY SIMILARITY.
FT DISULFID 757 768 BY SIMILARITY.
FT DISULFID 762 777 BY SIMILARITY.
FT DISULFID 779 788 BY SIMILARITY.
FT DISULFID 795 806 BY SIMILARITY.
FT DISULFID 800 815 BY SIMILARITY.
FT DISULFID 817 826 BY SIMILARITY.
FT DISULFID 833 844 BY SIMILARITY.
FT DISULFID 838 855 BY SIMILARITY.
FT DISULFID 857 866 BY SIMILARITY.
FT DISULFID 873 884 BY SIMILARITY.
FT DISULFID 878 893 BY SIMILARITY.
FT DISULFID 895 904 BY SIMILARITY.
FT DISULFID 911 922 BY SIMILARITY.
FT DISULFID 916 931 BY SIMILARITY.
FT DISULFID 933 942 BY SIMILARITY.
FT DISULFID 987 998 BY SIMILARITY.
FT DISULFID 992 1007 BY SIMILARITY.
FT DISULFID 1009 1018 BY SIMILARITY.
FT DISULFID 1025 1036 BY SIMILARITY.
FT DISULFID 1030 1045 BY SIMILARITY.
FT DISULFID 1047 1056 BY SIMILARITY.
FT DISULFID 1063 1074 BY SIMILARITY.
FT DISULFID 1068 1083 BY SIMILARITY.
FT DISULFID 1085 1094 BY SIMILARITY.
FT DISULFID 1101 1122 BY SIMILARITY.
FT DISULFID 1116 1131 BY SIMILARITY.
FT DISULFID 1133 1142 BY SIMILARITY.
FT DISULFID 1149 1160 BY SIMILARITY.
FT DISULFID 1154 1169 BY SIMILARITY.
FT DISULFID 1171 1180 BY SIMILARITY.
FT DISULFID 1187 1198 BY SIMILARITY.
FT DISULFID 1192 1207 BY SIMILARITY.
FT DISULFID 1209 1218 BY SIMILARITY.
FT DISULFID 1225 1244 BY SIMILARITY.
FT DISULFID 1238 1253 BY SIMILARITY.
FT DISULFID 1255 1264 BY SIMILARITY.
FT DISULFID 1271 1284 BY SIMILARITY.
FT DISULFID 1276 1293 BY SIMILARITY.
FT DISULFID 1295 1304 BY SIMILARITY.
FT DISULFID 1311 1322 BY SIMILARITY.
FT DISULFID 1316 1334 BY SIMILARITY.

```

Query Match 12.7% Score 74; DB 1; Length 2531;

Best Local Similarity 25.7%; Pred. No. 13; Mismatches 52; Indels 40; Gaps 8;

Matches 36; Conservative 12;

QY 4 LCLLLLPVL---GLVSSKTLCSMEEALNERIQEYAGSLIFRAISS--IGLEQCS-----54

Db 9 LCLTLLPALARGLRCSQPS---GTCINGGRCEVANGTCVCSGAFVGRQRCQDPSPCL 64

QY 54 -----VTSRGL-----ATCPRGFAVTC--TCGSAC-----GSDVRAETT 88
 DB 65 STPCNACTCYVDGGIVDYACSPGLGSLCLTPLANACLANPCRNCGTCDLLTLTE 124
 QY 89 CHQCAGMDWTGARCRCVOP 108
 DB 125 YKCRCP-PWGSKGSCQADP 143

RESULT 8

ID PRIA_BORBU STANDARD; PRT: 660 AA.
 AC Q45032; 051047;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DE PRIMOSOMAL PROTEIN N' (REPLICATION FACTOR Y).
 GN PRIA OR BB0014.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HB19;
 RA Boursaux-Eude C.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35210 / B31;
 RX MEDLINE; 98065943.
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
 RA Dougherty B., Tomb J.F., Fleischmann R.D., Richardson D.,
 RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
 RA van Vugt R., Palmer N., Adams M.D., Cocayne J.D., Weidman J.,
 RA Utterback T., Watthey L., McDonald L., Artach P., Bowman C.,
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
 RA Smith H.O., Venter J.C.;
 RA "Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi";
 RT Nature 390:580-586(1997).

CC -!- FUNCTION: RECOGNIZES A SPECIFIC HAIRPIN SEQUENCE ON PHIX SSDNA.
 CC THIS STRUCTURE IS THEN RECOGNIZED AND BOUND BY PROTEINS PRIB AND
 CC PRIC. FORMATION OF THE PRIMOSOME PROCEEDS WITH THE SUBSEQUENT
 CC ACTIONS OF DNAB, DNAC, DNAT AND PRIMASE. PRIA THEN FUNCTIONS AS A
 CC HELICASE WITHIN THE PRIMOSOME (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE PRIA SUBFAMILY OF HELICASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X97449; CA66080.1; -.
 DR EMBL; AE001115; AAC66393.1; -.
 DR TIGR; BB0014; -.
 DR PFAM; PF00271; Helicase_C; 1.
 DR DNPA; PF00271; Helicase; ATP-binding; Helicase; Primosome;
 KW Zinc-finger.
 FT NP_BIND 158 165 ATP (POTENTIAL).
 FT SITE 256 259 DEEH BOX.
 FT ZN_FING 370 382 C4-TYPE (POTENTIAL).
 FT ZN_FING 397 413 C4-TYPE (POTENTIAL).
 FT CONFLICT 126 126 P -> L (IN REF. 1).
 FT CONFLICT 555 555 D -> N (IN REF. 1).
 SQ SEQUENCE 660 AA; 77551 MW; 7FIAB19E04F0EFC3 CRC64;

Query Match 12.7%; Score 73.5; DB 1; Length 660;
 Best Local Similarity 29.2%; Pred. No. 4.2;
 Matches 28; Conservative 13; Mismatches 26; Indels 29; Gaps 7;

QY 16 VSKTKCSMEEAINEIRIQEVAGSLIFRAISSIG-----LEQSVTSRGDLATCPRGFAVTG 71
 DB 337 ISSELLYSIOKSLNEKRQ-----SLIF--INRGYLNLECNEC---GHICCP-----N 381
 QY 72 CTCG-----SAGCSWDVRAETTCCHC-QCAGMD 97
 DB 382 CSFGLIYHKKNKLLCHYCYKTKTASHCPQCESKD 417
 RESULT 9
 ID ADP1_YEAST STANDARD; PRT: 1049 AA.
 AC P25371;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DE PROBABLE ATP-DEPENDENT PERMEASE PRECURSOR.
 GN ADP1 OR YCR011C OR YCR11C OR YCR105.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 92160395.
 RA Purnelle B., Skala J., Goffeau A.;
 RT "The product of the YCR105 gene located on the chromosome III from
 RT Saccharomyces cerevisiae presents homologies to ATP-dependent
 RT permeases";
 RL Yeast 7:867-872(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 92327849.
 RA Skala J., Purnelle B., Goffeau A.;
 RT "The complete sequence of a 10.8 kb segment distal of SUF2 on the
 RT right arm of chromosome III from Saccharomyces cerevisiae reveals
 RT seven open reading frames including the RVS161, ADP1 and PKG genes";
 RL Yeast 8:409-417(1992).
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
 CC (ABC TRANSPORTERS). MDR SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X59720; CA442328.1; -.
 DR PIR; S19421; S19421.
 DR PIR; S40914; S40914.
 DR SGD; L000049; ADP1.
 DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
 DR PFAM; PF00005; ABC_tran; 1.
 KW ATP-binding; Transmembrane; Glycoprotein; Transport; Signal.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 1049 PROBABLE ATP-DEPENDENT PERMEASE.
 FT NP_BIND 423 430 ATP (BY SIMILARITY).
 FT TRANSMEM 325 345 POTENTIAL.
 FT TRANSMEM 464 481 POTENTIAL.
 FT TRANSMEM 794 814 POTENTIAL.
 FT TRANSMEM 829 849 POTENTIAL.
 FT TRANSMEM 878 898 POTENTIAL.
 FT TRANSMEM 910 930 POTENTIAL.
 FT TRANSMEM 938 958 POTENTIAL.
 FT TRANSMEM 1001 1021 POTENTIAL.
 FT TRANSMEM 1025 1045 POTENTIAL.
 FT CARBOHYD 50 50 POTENTIAL.
 FT CARBOHYD 114 114 POTENTIAL.
 FT CARBOHYD 165 165 POTENTIAL.
 FT CARBOHYD 221 221 POTENTIAL.

FT CARBOHYD 815 815 POTENTIAL.
 FT CARBOHYD 935 935 POTENTIAL.
 FT CARBOHYD 960 960 POTENTIAL.
 FT CARBOHYD 971 971 POTENTIAL.
 SQ SEQUENCE 1049 AA; 117231 MW; ABC9CE54BCFD6A3 CRC64;

Query Match 12.6%; Score 73; DB 1; Length 1049;
 Best Local Similarity 35.9%; Pred. No. 7.1;
 Matches 23; Conservative 4; Mismatches 21; Indels 16; Gaps 6;

QY 50 EQCVTSRGDLATCRPGFVAVTCT---CG-----SACGSWD--VRAET-TCHQCACGMDWT 99
 DB 71 ECNSYGR---CECIEGFAGDCLPLCGGLSPDESGNDRPIRAQNDTCHCD---NGWG 124
 QY 100 GARC 103
 DB 125 GINC 128

RESULT 10
 BAR3_CHITE STANDARD; PRT; 1700 AA.
 AC Q03376;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE BALBIANI RING PROTEIN 3 PRECURSOR.
 GN BR3.
 OS Chironomus tentans (Midge).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
 CC Chironomidae; Chironomidae; Chironominae; Chironomus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-SALIVARY GLAND;
 RX MEDLINE; 90172404.
 RA Paulsson G., Lendahl U., Galli J., Ericsson C., Wieslander L.;
 RT "The Balbiani ring 3 gene in Chironomus tentans has a diverged
 RT repetitive structure split by many introns.";
 RL J. Mol. Biol. 211:331-349(1990).
 CC -!- FUNCTION: USED BY THE LARVAE TO CONSTRUCT A SUPRAMOLECULAR
 CC STRUCTURE, THE LARVAL TUBE. BALBIANI RING PROTEIN 3 COULD PLAY A
 CC ROLE AS A TRANSPORT PROTEIN THAT BINDS TO OTHER PROTEINS
 CC INTRACELLULARLY AND IN THE GLAND LUMEN IN ORDER TO PREVENT THESE
 CC FROM FORMING WATER-INSOLUBLE FIBERS TOO EARLY.
 CC -!- SUBCELLULAR LOCATION: SECRETED.
 CC -!- TISSUE SPECIFICITY: SALIVARY GLAND.
 CC -!- DOMAIN: HAS 82 APPROXIMATE REPEATS OF CYS-X-CYS-X-CYS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X52263; CAA36506.1;
 DR PIR; S08167; S08167.
 DR HSSP; P18055; 2MRB.
 DR PRINTS; PR00876; MTNEMATODE.
 KW Repeat; Signal.
 FT SIGNAL
 FT CHAIN ? 1700 BALBIANI RING PROTEIN 3.
 SQ SEQUENCE 1700 AA; 186145 MW; 34202B28521B0815 CRC64;

Query Match 12.6%; Score 73; DB 1; Length 1700;
 Best Local Similarity 32.6%; Pred. No. 11;
 Matches 14; Conservative 7; Mismatches 12; Indels 10; Gaps 2;

QY 62 TCPRGFAV---TGCTC-----GSACGSWDVRAETTCQCA 94

DB 1508 TCKQGFSPKSGCKILECNKKDPGCGAKKINQCTCKCECA 1550
 RESULT 11
 NOTC_BRARE STANDARD; PRT; 2437 AA.
 AC P46530;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN PRECURSOR.
 GN NOTCH.
 OS Brachydanio rerio (Zebrafish) (Zebra danio).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
 CC Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
 CC Cyprinidae; Cyprinidae; Rasbora; Danio.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-EMBRYO;
 RX MEDLINE; 94128602.
 RA Bierkamp C., Campos-Ortega J.A.;
 RT "A zebrafish homologue of the Drosophila neurogenic gene Notch and
 RT its pattern of transcription during early embryogenesis.";
 RL Mech. Dev. 43:87-100(1993).
 CC -!- FUNCTION: IMPLICATED IN CELL FATE SPECIFICATIONS DURING
 CC EMBRYO DEVELOPMENT. MAY BE INVOLVED IN THE FORMATION OF THE
 CC NEURAL PLATE, NOTOCHORD AND BRAIN VESICLES.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN ALL CELLS IN PREGASTRULATION
 CC STAGES. DURING GASTRULATION IS DIFFERENTIALLY EXPRESSED,
 CC ACCUMULATING PREDOMINANTLY IN THE PRECHORDAL MESODERM AND
 CC NOTOCHORD. AT THE END OF GASTRULATION, EXPRESSED ALONG THE
 CC ANTERIOR-POSTERIOR AXIS INCLUDING THE DEVELOPING NEURAL PLATE
 CC AND DIFFERENTIATING MESODERM. ALSO PRESENT IN THE DEVELOPING
 CC BRAIN AND HEAD REGIONS.
 CC -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
 CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X69088; CAA48831.1;
 DR HSSP; P00740; 1IXA.
 DR ZFIN; ZDB-GENE-990415-173; NOTCH.
 DR PRINTS; PR00009; EGFTEF.
 DR PRINTS; PR00010; EGFBLD.
 DR PROSITE; PS00010; ASX_HYDROXYL; 23.
 DR PROSITE; PS00022; EGF_1; 34.
 DR PROSITE; PS01866; EGF_2; 28.
 DR PROSITE; PS01887; EGF_CA; 22.
 DR PFAM; PF00008; EGF; 36.
 DR PFAM; PF00023; ank; 6.
 DR PFAM; PF00066; notch; 3.
 KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Signal; Glycoprotein.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 2437 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN.
 FT DOMAIN 21 1724 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1725 1747 POTENTIAL.
 FT DOMAIN 1748 2437 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 21 57 EGF-LIKE 1.
 FT DOMAIN 58 98 EGF-LIKE 2.
 FT DOMAIN 101 138 EGF-LIKE 3.
 FT DOMAIN 139 175 EGF-LIKE 4.
 FT DOMAIN 177 215 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 177 215

```
FT DOMAIN 217 254 EGF-LIKE 6. 415 428 BY SIMILARITY.
FT DOMAIN 256 292 EGF-LIKE 7. 422 437 BY SIMILARITY.
FT DOMAIN 294 332 EGF-LIKE 8. 439 448 BY SIMILARITY.
FT DOMAIN 334 370 EGF-LIKE 9. 455 466 BY SIMILARITY.
FT DOMAIN 371 409 EGF-LIKE 10. 460 475 BY SIMILARITY.
FT DOMAIN 411 449 EGF-LIKE 11. 477 486 BY SIMILARITY.
FT DOMAIN 451 487 EGF-LIKE 12. 493 503 BY SIMILARITY.
FT DOMAIN 489 524 EGF-LIKE 13. 498 512 BY SIMILARITY.
FT DOMAIN 526 562 EGF-LIKE 14. 514 523 BY SIMILARITY.
FT DOMAIN 564 599 EGF-LIKE 15. 530 541 BY SIMILARITY.
FT DOMAIN 601 637 EGF-LIKE 16. 541 550 BY SIMILARITY.
FT DOMAIN 639 674 EGF-LIKE 17. 552 561 BY SIMILARITY.
FT DOMAIN 676 712 EGF-LIKE 18. 568 578 BY SIMILARITY.
FT DOMAIN 714 749 EGF-LIKE 19. 573 587 BY SIMILARITY.
FT DOMAIN 751 787 EGF-LIKE 20. 589 598 BY SIMILARITY.
FT DOMAIN 789 825 EGF-LIKE 21. 605 616 BY SIMILARITY.
FT DOMAIN 827 865 EGF-LIKE 22. 616 625 BY SIMILARITY.
FT DOMAIN 867 903 EGF-LIKE 23. 627 636 BY SIMILARITY.
FT DOMAIN 905 941 EGF-LIKE 24. 643 653 BY SIMILARITY.
FT DOMAIN 943 979 EGF-LIKE 25. 648 662 BY SIMILARITY.
FT DOMAIN 981 1017 EGF-LIKE 26. 664 673 BY SIMILARITY.
FT DOMAIN 1019 1055 EGF-LIKE 27. 680 691 BY SIMILARITY.
FT DOMAIN 1057 1093 EGF-LIKE 28. 691 700 BY SIMILARITY.
FT DOMAIN 1095 1141 EGF-LIKE 29. 702 711 BY SIMILARITY.
FT DOMAIN 1143 1179 EGF-LIKE 30. 718 728 BY SIMILARITY.
FT DOMAIN 1181 1217 EGF-LIKE 31. 723 737 BY SIMILARITY.
FT DOMAIN 1219 1263 EGF-LIKE 32. 739 748 BY SIMILARITY.
FT DOMAIN 1265 1303 EGF-LIKE 33. 755 766 BY SIMILARITY.
FT DOMAIN 1305 1344 EGF-LIKE 34. 766 775 BY SIMILARITY.
FT DOMAIN 1346 1382 EGF-LIKE 35. 777 786 BY SIMILARITY.
FT DOMAIN 1385 1423 EGF-LIKE 36. 793 804 BY SIMILARITY.
FT DOMAIN 1446 1561 3 X LIN/NOTCH REPEATS. 798 813 BY SIMILARITY.
FT REPEAT 1487 1520 LIN/NOTCH 1. 815 824 BY SIMILARITY.
FT REPEAT 1521 1561 LIN/NOTCH 2. 831 842 BY SIMILARITY.
FT DOMAIN 1561 2074 6 X ANK MOTIF REPEATS. 836 853 BY SIMILARITY.
FT REPEAT 1861 1891 ANK MOTIF 1. 855 864 BY SIMILARITY.
FT REPEAT 1892 1940 ANK MOTIF 1. 871 882 BY SIMILARITY.
FT REPEAT 1941 1974 ANK MOTIF 1. 876 891 BY SIMILARITY.
FT REPEAT 1975 2007 ANK MOTIF 1. 893 902 BY SIMILARITY.
FT REPEAT 2008 2040 ANK MOTIF 1. 909 920 BY SIMILARITY.
FT REPEAT 2041 2074 ANK MOTIF 1. 914 929 BY SIMILARITY.
FT DOMAIN 2075 2276 POLY-GLN (OPA-REPEAT). 931 940 BY SIMILARITY.
FT DISULFID 25 35 BY SIMILARITY. 947 958 BY SIMILARITY.
FT DISULFID 29 45 BY SIMILARITY. 952 967 BY SIMILARITY.
FT DISULFID 47 56 BY SIMILARITY. 969 978 BY SIMILARITY.
FT DISULFID 62 73 BY SIMILARITY. 1023 1034 BY SIMILARITY.
FT DISULFID 67 86 BY SIMILARITY. 1028 1043 BY SIMILARITY.
FT DISULFID 88 97 BY SIMILARITY. 1045 1054 BY SIMILARITY.
FT DISULFID 105 116 BY SIMILARITY. 1061 1072 BY SIMILARITY.
FT DISULFID 110 126 BY SIMILARITY. 1066 1081 BY SIMILARITY.
FT DISULFID 128 137 BY SIMILARITY. 1083 1092 BY SIMILARITY.
FT DISULFID 143 154 BY SIMILARITY. 1099 1120 BY SIMILARITY.
FT DISULFID 148 163 BY SIMILARITY. 1114 1129 BY SIMILARITY.
FT DISULFID 165 174 BY SIMILARITY. 1131 1140 BY SIMILARITY.
FT DISULFID 181 194 BY SIMILARITY. 1147 1158 BY SIMILARITY.
FT DISULFID 188 203 BY SIMILARITY. 1152 1167 BY SIMILARITY.
FT DISULFID 205 214 BY SIMILARITY. 1169 1178 BY SIMILARITY.
FT DISULFID 221 232 BY SIMILARITY. 1185 1196 BY SIMILARITY.
FT DISULFID 226 242 BY SIMILARITY. 1190 1205 BY SIMILARITY.
FT DISULFID 244 253 BY SIMILARITY. 1207 1216 BY SIMILARITY.
FT DISULFID 260 271 BY SIMILARITY.
FT DISULFID 265 280 BY SIMILARITY.
FT DISULFID 282 291 BY SIMILARITY.
FT DISULFID 298 311 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 322 331 BY SIMILARITY.
FT DISULFID 338 349 BY SIMILARITY.
FT DISULFID 343 358 BY SIMILARITY.
FT DISULFID 360 369 BY SIMILARITY.
FT DISULFID 375 386 BY SIMILARITY.
FT DISULFID 380 397 BY SIMILARITY.
FT DISULFID 399 408 BY SIMILARITY.
```

Query Match 12.6%; Score 73; DB 1; Length 2437;

Best Local Similarity 30.6%; Pred. No. 15;

Matches 22; Conservative 3; Mismatches 21; Indels 26; Gaps 5;

QY 51 CQSVTSRGD-----LATCPRGFAVTGCTC-----GSACGSWDVRAETTC 91

Db 1028 CQGGSCQDGVGYTKCTCPHGY--TGLNCQSLVRWCDSPOKNGGSC--WQOGASFTQC 1083

QY 92 QCAGMDWTGARC 103

Db 1084 ---ASGWTGIYC 1092

RESULT 12
 LMG2_HUMAN STANDARD; PRT; 1193 AA.
 AC Q13753: Q14941; Q02536; Q02537;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 15-FEB-2000 (Rel. 39, Last annotation update)
 DE LAMININ GAMMA-2 CHAIN PRECURSOR.
 GN LAMC2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RX MEDLINE: 96230326.
 RA Alerenne T., Haakana H., Sainio K., Kallunki T., Kallunki P.,
 RA Sariola H., Tryggvason K.;
 RT "Structure of the human laminin gamma 2 chain gene (LAMC2):
 RT alternative splicing with different tissue distribution of two
 RT transcripts.";
 RL Genomics 32:54-64(1996).
 RN [2]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1090-1114.
 RC TISSUE=EPIDERMIS, AND KERATINOCYTES;
 RX MEDLINE: 94139694.
 RA Vaillay J., Verrando P., Champilaud M.F., Gerecke D., Wagman D.W.,
 RA Baudoin C., Aberdam D., Burgeson R., Bauer E., Ortonne J.P.;
 RT "The 100-kDa chain of nectin/kalinin is a laminin B2 chain variant.";
 RT Eur. J. Biochem. 219:209-218(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=FIBROSARCOMA;
 RX MEDLINE: 93016279.
 RA Kallunki P., Sainio K., Eddy R., Byers M., Kallunki T.,
 RA Sariola H., Beck K., Hirvonen H., Shows T.B., Tryggvason K.;
 RT "A truncated laminin chain homologous to the B2 chain: structure,
 RT spatial expression, and chromosomal assignment.";
 RL J. Cell Biol. 119:679-693(1992).
 CC -!- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
 CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
 CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
 CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
 CC -!- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
 CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
 CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
 CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
 CC THE GAMMA-2 CHAIN IS A SUBUNIT OF LAMININ-5 (EPILIGRIN/KALININ/
 CC NICEIN).
 CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
 CC MEMBRANES (MAJOR COMPONENT).
 CC -!- ALTERNATIVE PRODUCTS: TWO ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- TISSUE SPECIFICITY: THE LARGE VARIANT IS EXPRESSED ONLY IN
 CC SPECIFIC EPITHELIAL CELLS OF EMBRYONIC AND NEONATAL TISSUES. IN
 CC 17-WEEK OLD EMBRYO THE SMALL VARIANT IS FOUND IN CEREBRAL CORTEX,
 CC LUNG, AND DISTAL TUBES OF KIDNEY, BUT NOT IN EPITHELIA EXCEPT FOR
 CC DISTAL TUBULI.
 CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -!- DOMAIN: DOMAIN IV IS GLOBULAR.
 CC -!- DISEASE: JUNCTIONAL EPIDERMOLYSIS BULLOSA GRAVIS IS A BLISTERING
 CC DISORDER IN SKIN THAT IS CHARACTERIZED BY A SEPARATION OF BASAL
 CC CELLS FROM THE BASEMENT MEMBRANE DUE TO A DECREASED NUMBER OF
 CC HEMIDESMOSOMES. LAMININ-5 IS MISSING FROM THE BASEMENT MEMBRANE OF
 CC PATIENTS WITH THE GRAVIS FORM OF EPIDERMOLYSIS BULLOSA.
 CC -!- SIMILARITY: CONTAINS 7.5 LAMININ EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
 or send an email to license@isb-sib.ch).

 CC EMBL; U31201; AAC50457.1; -
 CC EMBL; U31178; AAC50457.1; JOINED.
 CC EMBL; U31179; AAC50457.1; JOINED.
 CC EMBL; U31180; AAC50457.1; JOINED.
 CC EMBL; U31181; AAC50457.1; JOINED.
 CC EMBL; U31182; AAC50457.1; JOINED.
 CC EMBL; U31183; AAC50457.1; JOINED.
 CC EMBL; U31184; AAC50457.1; JOINED.
 CC EMBL; U31186; AAC50457.1; JOINED.
 CC EMBL; U31187; AAC50457.1; JOINED.
 CC EMBL; U31188; AAC50457.1; JOINED.
 CC EMBL; U31189; AAC50457.1; JOINED.
 CC EMBL; U31190; AAC50457.1; JOINED.
 CC EMBL; U31191; AAC50457.1; JOINED.
 CC EMBL; U31192; AAC50457.1; JOINED.
 CC EMBL; U31193; AAC50457.1; JOINED.
 CC EMBL; U31194; AAC50457.1; JOINED.
 CC EMBL; U31195; AAC50457.1; JOINED.
 CC EMBL; U31196; AAC50457.1; JOINED.
 CC EMBL; U31197; AAC50457.1; JOINED.
 CC EMBL; U31198; AAC50457.1; JOINED.
 CC EMBL; U31199; AAC50457.1; JOINED.
 CC EMBL; U31200; AAC50456.1; -
 CC EMBL; U31178; AAC50456.1; JOINED.
 CC EMBL; U31179; AAC50456.1; JOINED.
 CC EMBL; U31180; AAC50456.1; JOINED.
 CC EMBL; U31181; AAC50456.1; JOINED.
 CC EMBL; U31182; AAC50456.1; JOINED.
 CC EMBL; U31183; AAC50456.1; JOINED.
 CC EMBL; U31184; AAC50456.1; JOINED.
 CC EMBL; U31186; AAC50456.1; JOINED.
 CC EMBL; U31187; AAC50456.1; JOINED.
 CC EMBL; U31188; AAC50456.1; JOINED.
 CC EMBL; U31189; AAC50456.1; JOINED.
 CC EMBL; U31190; AAC50456.1; JOINED.
 CC EMBL; U31191; AAC50456.1; JOINED.
 CC EMBL; U31192; AAC50456.1; JOINED.
 CC EMBL; U31193; AAC50456.1; JOINED.
 CC EMBL; U31194; AAC50456.1; JOINED.
 CC EMBL; U31195; AAC50456.1; JOINED.
 CC EMBL; U31196; AAC50456.1; JOINED.
 CC EMBL; U31197; AAC50456.1; JOINED.
 CC EMBL; U31198; AAC50456.1; JOINED.
 CC EMBL; X73902; CAA52108.1; -
 CC EMBL; Z15008; CAA78728.1; -
 CC EMBL; Z15009; CAA78729.1; -
 CC HSP; P02468; ITLE.
 CC MIN; 150292; -
 CC PROSITE; PS00022; EGF_1; 4.
 CC PROSITE; PS01186; EGF_2; 2.
 CC PROSITE; PS01248; LAMININ_TYPE_EGF; 6.
 CC PFAM; PF00052; laminin_B; 1.
 CC PFAM; PF00053; laminin_EGF; 6.
 CC Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 CC Laminin EGF-like domain; Cell adhesion; Repeat; Signal;
 CC Alternative splicing.
 CC SIGNAL 1 21
 CC CHAIN 22 1193
 CC DOMAIN 28 196
 CC
 CC DOMAIN 28 83
 CC DOMAIN 84 130
 CC DOMAIN 139 186
 CC DOMAIN 187 196
 CC DOMAIN 197 381
 CC DOMAIN 382 602
 CC DOMAIN 382 415
 CC DOMAIN 416 461
 CC
 CC POTENTIAL.
 CC LAMININ GAMMA-2 CHAIN.
 CC 3.5 X LAMININ EGF-LIKE REPEATS (DOMAIN
 CC V).
 CC LAMININ EGF-LIKE 1.
 CC LAMININ EGF-LIKE 2.
 CC LAMININ EGF-LIKE 3.
 CC LAMININ EGF-LIKE 4 (N-TERMINAL).
 CC LAMININ DOMAIN IV.
 CC 4 X LAMININ EGF-LIKE REPEATS (DOMAIN V).
 CC LAMININ EGF-LIKE 4 (C-TERMINAL).
 CC LAMININ EGF-LIKE 5.


```
FT DOMAIN 1037 1083 EGF-LIKE 27.
FT DOMAIN 1085 1121 EGF-LIKE 28.
FT DOMAIN 1123 1159 EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1161 1204 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1206 1245 EGF-LIKE 31.
FT DOMAIN 1247 1288 EGF-LIKE 32.
FT DOMAIN 1290 1326 EGF-LIKE 33.
FT DOMAIN 1336 1374 EGF-LIKE 34.
FT REPEAT 1388 1428 LIN/NOTCH 1.
FT REPEAT 1429 1467 LIN/NOTCH 2.
FT REPEAT 1468 1503 LIN/NOTCH 3.
FT REPEAT 1784 1816 CDC10/SWI6 1.
FT REPEAT 1817 1865 CDC10/SWI6 2.
FT REPEAT 1866 1898 CDC10/SWI6 3.
FT REPEAT 1899 1932 CDC10/SWI6 4.
FT REPEAT 1933 1965 CDC10/SWI6 5.
FT REPEAT 1966 1998 CDC10/SWI6 6.
FT DISULFID 43 55 BY SIMILARITY.
FT DISULFID 49 66 BY SIMILARITY.
FT DISULFID 68 77 BY SIMILARITY.
FT DISULFID 83 94 BY SIMILARITY.
FT DISULFID 88 107 BY SIMILARITY.
FT DISULFID 109 118 BY SIMILARITY.
FT DISULFID 124 135 BY SIMILARITY.
FT DISULFID 129 145 BY SIMILARITY.
FT DISULFID 147 156 BY SIMILARITY.
FT DISULFID 163 175 BY SIMILARITY.
FT DISULFID 189 194 BY SIMILARITY.
FT DISULFID 186 195 BY SIMILARITY.
FT DISULFID 202 213 BY SIMILARITY.
FT DISULFID 207 223 BY SIMILARITY.
FT DISULFID 225 234 BY SIMILARITY.
FT DISULFID 241 252 BY SIMILARITY.
FT DISULFID 246 261 BY SIMILARITY.
FT DISULFID 263 272 BY SIMILARITY.
FT DISULFID 279 292 BY SIMILARITY.
FT DISULFID 286 301 BY SIMILARITY.
FT DISULFID 303 312 BY SIMILARITY.
FT DISULFID 319 330 BY SIMILARITY.
FT DISULFID 324 339 BY SIMILARITY.
FT DISULFID 341 350 BY SIMILARITY.
FT DISULFID 356 367 BY SIMILARITY.
FT DISULFID 361 378 BY SIMILARITY.
FT DISULFID 380 389 BY SIMILARITY.
FT DISULFID 396 409 BY SIMILARITY.
FT DISULFID 403 418 BY SIMILARITY.
FT DISULFID 420 429 BY SIMILARITY.
FT DISULFID 436 447 BY SIMILARITY.
FT DISULFID 441 456 BY SIMILARITY.
FT DISULFID 458 467 BY SIMILARITY.
FT DISULFID 474 485 BY SIMILARITY.
FT DISULFID 479 494 BY SIMILARITY.
FT DISULFID 496 505 BY SIMILARITY.
FT DISULFID 512 523 BY SIMILARITY.
FT DISULFID 517 532 BY SIMILARITY.
FT DISULFID 534 543 BY SIMILARITY.
FT DISULFID 550 560 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 571 580 BY SIMILARITY.
FT DISULFID 587 598 BY SIMILARITY.
FT DISULFID 607 618 BY SIMILARITY.
FT DISULFID 609 618 BY SIMILARITY.
FT DISULFID 625 635 BY SIMILARITY.
FT DISULFID 630 644 BY SIMILARITY.
FT DISULFID 646 655 BY SIMILARITY.
FT DISULFID 662 673 BY SIMILARITY.
FT DISULFID 667 682 BY SIMILARITY.
FT DISULFID 684 693 BY SIMILARITY.
FT DISULFID 700 710 BY SIMILARITY.
FT DISULFID 705 719 BY SIMILARITY.
FT DISULFID 721 730 BY SIMILARITY.
FT DISULFID 739 750 BY SIMILARITY.
FT DISULFID 744 759 BY SIMILARITY.
```

```
FT DISULFID 761 770 BY SIMILARITY.
FT DISULFID 776 787 BY SIMILARITY.
FT DISULFID 781 797 BY SIMILARITY.
FT DISULFID 799 808 BY SIMILARITY.
FT DISULFID 815 827 BY SIMILARITY.
FT DISULFID 821 836 BY SIMILARITY.
FT DISULFID 838 847 BY SIMILARITY.
FT DISULFID 854 865 BY SIMILARITY.
FT DISULFID 859 874 BY SIMILARITY.
FT DISULFID 876 885 BY SIMILARITY.
FT DISULFID 892 902 BY SIMILARITY.
FT DISULFID 897 911 BY SIMILARITY.
FT DISULFID 913 922 BY SIMILARITY.
FT DISULFID 929 940 BY SIMILARITY.
FT DISULFID 934 949 BY SIMILARITY.
FT DISULFID 951 960 BY SIMILARITY.
FT DISULFID 967 978 BY SIMILARITY.
FT DISULFID 972 987 BY SIMILARITY.
FT DISULFID 989 998 BY SIMILARITY.
FT DISULFID 1005 1016 BY SIMILARITY.
FT DISULFID 1010 1023 BY SIMILARITY.
FT DISULFID 1025 1034 BY SIMILARITY.
FT DISULFID 1041 1062 BY SIMILARITY.
FT DISULFID 1056 1071 BY SIMILARITY.
FT DISULFID 1073 1082 BY SIMILARITY.
FT DISULFID 1089 1100 BY SIMILARITY.
FT DISULFID 1094 1109 BY SIMILARITY.
FT DISULFID 1111 1120 BY SIMILARITY.
FT DISULFID 1127 1138 BY SIMILARITY.
FT DISULFID 1132 1147 BY SIMILARITY.
FT DISULFID 1149 1158 BY SIMILARITY.
FT DISULFID 1165 1183 BY SIMILARITY.
FT DISULFID 1177 1192 BY SIMILARITY.
FT DISULFID 1194 1203 BY SIMILARITY.
FT DISULFID 1210 1223 BY SIMILARITY.
FT DISULFID 1215 1233 BY SIMILARITY.
FT DISULFID 1235 1244 BY SIMILARITY.
FT DISULFID 1251 1262 BY SIMILARITY.
FT DISULFID 1256 1276 BY SIMILARITY.
FT DISULFID 1278 1287 BY SIMILARITY.
FT DISULFID 1294 1305 BY SIMILARITY.
FT DISULFID 1299 1314 BY SIMILARITY.
FT DISULFID 1316 1325 BY SIMILARITY.
FT DISULFID 1340 1351 BY SIMILARITY.
FT DISULFID 1345 1362 BY SIMILARITY.
```

Query Match 12.4%; Score 72; DB 1; Length 2318;

Best Local Similarity 32.0%; Pred. No. 18;

Matches 24; Conservative 7; Mismatches 24; Indels 20; Gaps 6;

QY 47 IGLECQSVTSRGDIATCPRGFA-----VTGCTCGSA-C---GS-WDVRAETT 88

Db 1301 VGIPQQ-TARGPRCACPGLSGSPSCNRVSRASPSGATNASCAPCLHGSGCLPVQSVPF 1359

QY 89 CHQCACGMDWTGARC 103

Db 1360 FRCVCA-PGNGGPC 1373

RESULT 14

TYR2_HUMAN STANDARD; PRT; 519 AA.

ID TYR2_HUMAN

AC P40126;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 15-FEB-2000 (Rel. 39, Last annotation update)

DE DOPACHROME TAUOMERASE PRECURSOR (EC 5.3.3.12) (DT) (DCT) (DOPACHROME

DE DELTA-ISOMERASE) (TYROSINASE-RELATED PROTEIN 2) (TRP-2) (TRP2).

GN DCT OR TRP2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Primates; Catarrhini; Homnidae; Homo.

RN [1]

Search completed: May 26, 2000, 04:53:06
Job time: 6674 sec

Result No.	Query			ID	Description
	Score	Match	Length		
1	86	14.8	1059	5 P90884	P90884 caenorhabdi
2	84.5	14.5	1476	13 Q90285	Q90285 carassius
3	82.5	14.2	762	13 Q42373	Q42373 brachydani
4	82.5	14.2	1574	11 Q88281	Q88281 rattus nor
5	82	14.1	2321	4 Q9Y6L8	Q9Y6L8 homo sapien
6	78	13.4	594	5 Q9Y151	Q9Y151 drosophila
7	78	13.4	1106	5 Q17494	Q17494 caenorhabdi
8	76.5	13.2	589	11 Q88671	Q88671 rattus nor
9	76	13.1	2447	13 Q13149	Q13149 fugu rubri
10	75.5	13.0	153	4 Q75095	Q75095 homo sapien
11	75.5	13.0	387	11 Q06007	Q06007 mus muscul
12	74.5	12.8	75	5 Q96388	Q96388 perna virid
13	74.5	12.8	585	11 Q35675	Q35675 mus muscul
14	74.5	12.8	592	11 Q88516	Q88516 mus muscul
15	74.5	12.8	1698	5 Q94438	Q94438 chironomus
16	73.5	12.7	289	5 P91237	P91237 caenorhabdi
17	73.5	12.7	830	4 Q43701	Q43701 homo sapien
18	73.5	12.7	830	4 Q4162	Q4162 homo sapien
19	71.5	12.3	105	5 Q17187	Q17187 bombyx mori
20	71.5	12.3	237	4 Q75767	Q75767 homo sapien

QY	20	103MEATNE


```

KW  Glycoprotein. 1
FT  NON_TER 762
FT  NON_TER 762
SQ  SEQUENCE 762 AA; 83560 MW; 9FE1F01E CRC32;

Query Match 14.2%; Score 82.5; DB 13; Length 762;
Best Local Similarity 32.1%; Pred. No. 0.5;
Matches 27; Conservative 9; Mismatches 31; Indels 17; Gaps 5;

QY 36 AGSLIFRAISIGLEQC-----SVTSRGDIATCPRGFAVTCGT-----CGS-ACGSWDV 83
    ||| :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 142 AGSNCRSMCKCLPCYNGGSCLTITTRGARTCIQGGPLCQHRSDNGCSSKPCHHGGL 201
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 84 RAETTC-----HCQAGMDWTGARC 103
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 202 CTEETSYPFFHCQCTN-GWKGKRC 224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
ID O88281 PRELIMINARY; PRT; 1574 AA.
AC O88281;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE MEGF6.
DE MEGF6.
GN MEGF6.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
[1]
SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-BRAIN;
RX MEDLINE; 98360089.
RA NAKAJIYAMA M., NAKAJIYAMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT "Identification of high-molecular-weight proteins with multiple EGF-
RT like motifs by motif-trap screening.";
RL Genomics 51:27-34(1998).
DR EMBL; AB011532; BAA32462.1; -.
DR HSP; P00736; IAPQ.
DR PROSITE; PS00010; ASX-HYDROXYL; 5.
DR PROSITE; PS01187; EGF_CA; 5.
DR PFAM; PF00008; EGF; 25.
KW Glycoprotein; EGF-like domain.
SQ SEQUENCE 1574 AA; 165445 MW; C49E6EA6 CRC32;

Query Match 14.2%; Score 82.5; DB 11; Length 1574;
Best Local Similarity 35.6%; Pred. No. 1.1;
Matches 21; Conservative 3; Mismatches 26; Indels 9; Gaps 2;

QY 47 IGLSCQSVTSRGDIATCPRGFAVTCGTGCGSACGSWDVRAETTCQCQAGMDWTGRCR 105
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 813 VGSRCQD-----TCSAGWYGTGQTRCACANDGCDPTTGRCSCA-PGWTGLSCQR 862
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
QY9Y6L8
ID QY9Y6L8 PRELIMINARY; PRT; 2321 AA.
AC QY9Y6L8;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE NOTCH3.
DE NOTCH3.
GN NOTCH3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
[1]
SEQUENCE FROM N.A.
RA GUNEL M., ARTAVANIS-TSAKONAS S.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.

```

```

Db      182  NKSCDHITGEILCRT-GYGLTCEH-----PCPAGLYGPGCKLKCNCHEGECNHVT 232
QY      89  CHQCQAGMDWTGARC 103
        |||  |||:  |
Db      233  GQCOCL-PGWTGNC 246
        |||  |||:  |

RESULT 7
QID17494 PRELIMINARY; PRT; 1106 AA.
AC Q17494;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE B0393.5 PROTEIN.
GN B0393.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidae; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RA SULSTON J.;
RP SEQUENCE FROM N.A.
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RA SULSTON J.;
RP SEQUENCE FROM N.A.
RL MEDLINE: 94150718.
RX WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMAILDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; 237983; CAA86058.1; -.
DR HSSP; P07204; 2ADX.
DR PROSITE; PS00010; ASX_HYDROXYL; 6.
DR PROSITE; PS01187; EGF_CA; 5.
DR PFAM; PF00008; EGF; 5.
DR Glycoprotein; EGF-like domain.
KW SEQUENCE 1106 AA; 121137 MW; 2C93CB51 CRC32;
SQ

Query Match 13.4%; Score 78; DB 5; Length 1106;
Best Local Similarity 31.1%; Pred.No.2.3;
Matches 19; Conservative 4; Mismatches 22; Indels 16; Gaps 3;

QY 59 DLATCPRG-FAYTGTCGSAGSDVRAETTCCHC-----QCAG----MDWTGAR 102
   |||||  |||  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
Ddb 772 DLTQCPPLGRLGDLCDRECAAGHYGNCSTCHDGSVACDVTGCPALCRAGWEGSS 831
   |||||  |||  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
QY 103 C 103
   |
Ddb 832 C 832

RESULT 8
QID88671 PRELIMINARY; PRT; 589 AA.
AC Q88671;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE DELTA 3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]

```

```

GN MG6F6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
[!]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 98360089.
RA NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT "Identification of high-molecular-weight proteins with multiple EGF-
ET like motifs by motif-trap screening.";
RL Genomics 51:27-34(1998).
DR EMBL; AB011539; BAA32467.1; -.
DR HSP; P02468; IKLO.
DR PFAM; PF00008; EGF; 2.
FT NON_TER 1
SQ SEQUENCE 153 AA; 15579 MW; 1AC64E20 CRC32;

Query Match 13.0%; Score 75.5; DB 4; Length 153;
Best Local Similarity 35.7%; Pred. No. 0.54;
Matches 20; Conservative 1; Mismatches 26; Indels 9; Gaps

QY 48 GLEQCSTVSRGDLATCPRGFAVTGCTCGSACSVDYRATTTCCHCAGMDWTGARC 103
| | | | | | | | : | | | | | | | | | | | | |
Db 51 GAPCDPVTG---LCLPPG---RSGATCNLCRRQGFGPSCTLHDCGG----GADC 97

RESULT 11
Q06007 PRELIMINARY; PRT; 387 AA.
AC Q06007;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (T-EMBLrel. 12, Last annotation update)
DE NOTCH1 PROTEIN HOMOLOG 1 (NOTCH A PROTEIN) (FRAGMENT).
GN NOTCH1 OR NOTCH A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [!]
RP SEQUENCE FROM N.A.
RC STRAIN-F1 (CBA X C57BL); TISSUE-WHOLE EMBRYO;
RX MEDLINE; 93178563.
RA LARDELLI M., LENDAHL U.;
RT "Notch A and notch B--two mouse Notch homologues coexpressed in a wide
variety of tissues.";
RL Exp. Cell Res. 204:364-372(1993).
DR EMBL; X68278; CAA48339.1; -.
DR HSP; P00743; LAPO.
DR MGD; MGI:97363; Notch1.
DR PFAM; PF00008; EGF; 6.
DR PFAM; PF00086; notch; 3.
KW Differentiation; Neurogenesis; Repeat.
FT NON_TER 1
FT NON_TER 387
SQ SEQUENCE 387 AA; 41497 MW; D1FD6C00 CRC32;

Query Match 13.0%; Score 75.5; DB 11; Length 387;
Best Local Similarity 29.1%; Pred. No. 1.4;
Matches 25; Conservative 4; Mismatches 24; Indels 33; Gaps

QY 48 GLEQCSTVSRGDLATCPRGFAVTGCTCGSACSVDYRATTTCCHCAGMDWTGARC 103
| | | | | | | | : | | | | | | | | | | | | |
Db 141 GRRCESVINGCRKPKNGGVCAVSNTARGFCRCPAGE--EGATCENDARTCGSLRCL 198
| | | | | | | | : | | | | | | | | | | | | |

QY 85 AETTCT-----HCQAGMDWTGARC 103
| | | | | | | | : | | | | | | | | | | | | |
Db 199 NGGTCSIGPRSPCTCLCLG-SFTGPEC 223
| | | | | | | | : | | | | | | | | | | | | |

RESULT 12

```

```
O96388
ID O96388 PRELIMINARY; PRT; 75 AA.
AC O96388;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE METALLOTHIONEIN.
OS Perna viridis.
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;
OC Mytilidae; Perna.
RN [1]
RP SEQUENCE FROM N.A.
RA KHOO H.W., WONG Y.W., PATEL K.H.;
RT "Green mussel (Perna viridis) metallothionein cDNA.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF036904; AD02054.1; -
SQ SEQUENCE 75 AA; 7415 MW; 942586D8 CRC32;

Query Match 12.8%; Score 74.5; DB 5; Length 75;
Best Local Similarity 27.9%; Pred. No. 0.33;
Matches 17; Conservative 9; Mismatches 18; Indels 17; Gaps 4;

QY 51 CQSVTRGDIATCPRGFAVTCGCSAGSDVRAETTCCHCQAGMDWTGARC---CRVQ 107
| : : : : | | | | : : | | | | : : | | | |
Db 5 CNIETQ--VCIGTGCGBECRCGDAC-----KCSGCGCGCSG-----CKVCKCQ 50
QY 108 P 108
Db 51 P 51

RESULT 13
ID O35675 PRELIMINARY; PRT; 585 AA.
AC O35675;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE M-DELTA-LIKE 3 GENE PRECURSOR.
GN DLL3 OR M-DELTA-LIKE 3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57B16 X DBA; TISSUE=PRIMITIVE STREAK;
RX MEDLINE; 97417575.
RA DUNWOODIE S.L., HENRIQUE D.M.P., HARRISON S.M., BEDDINGTON R.S.P.;
RT "Mouse Dll3: a novel divergent Delta gene which may complement the
function of other Delta homologues during early pattern formation in
the mouse embryo.";
RL Development 124:3065-3076(1997).
DR EMBL: Y11895; CAA72637.1; -
DR HSP; P00740; LIXA.
DR MGD; MGI:1096877; D113.
DR PFAM; PF00008; EGF; 6.
DR PRINTS; PR00010; EGFBL00D.
KW Signal; Glycoprotein.
FT SIGNAL 1 32 POTENTIAL.
SQ SEQUENCE 585 AA; 61129 MW; 0F5E6187 CRC32;

Query Match 12.8%; Score 74.5; DB 11; Length 585;
Best Local Similarity 28.8%; Pred. No. 2.8;
Matches 34; Conservative 8; Mismatches 43; Indels 33; Gaps 8;

QY 21 LC----SMEAINERIQEVA--GSLIFRAISSIGLEQ-----SVTSRGDLATCPRGF-- 68
| | : | : | : | : | : | : | : | : | : | : | : |
Db 245 LCTVPVSTSSCLNSRVPGPASTGCLLPGPCDGNPCANGSCSETSGSFECACPRGYG 304
DE DELTA-LIKE 3 GENE PRECURSOR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57B16 X DBA; TISSUE=PRIMITIVE STREAK;
RX MEDLINE; 97417575.
RA DUNWOODIE S.L., HENRIQUE D.M.P., HARRISON S.M., BEDDINGTON R.S.P.;
RT "Mouse Dll3: a novel divergent Delta gene which may complement the
function of other Delta homologues during early pattern formation in
the mouse embryo.";
RL Development 124:3065-3076(1997).
DR EMBL: Y11895; CAA72637.1; -
DR HSP; P00740; LIXA.
DR MGD; MGI:1096877; D113.
DR PFAM; PF00008; EGF; 6.
DR PRINTS; PR00010; EGFBL00D.
KW Signal; Glycoprotein.
FT SIGNAL 1 32 POTENTIAL.
SQ SEQUENCE 585 AA; 61129 MW; 0F5E6187 CRC32;

Query Match 12.8%; Score 74.5; DB 11; Length 592;
Best Local Similarity 28.8%; Pred. No. 2.9;
Matches 34; Conservative 8; Mismatches 43; Indels 33; Gaps 8;

QY 21 LC----SMEAINERIQEVA--GSLIFRAISSIGLEQ-----SVTSRGDLATCPRGF-- 68
| | : | : | : | : | : | : | : | : | : | : | : |
Db 245 LCTVPVSTSSCLNSRVPGPASTGCLLPGPCDGNPCANGSCSETSGSFECACPRGYG 304
DE DELTA-LIKE 3 GENE PRECURSOR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57B16 X DBA; TISSUE=PRIMITIVE STREAK;
RX MEDLINE; 97417575.
RA DUNWOODIE S.L., HENRIQUE D.M.P., HARRISON S.M., BEDDINGTON R.S.P.;
RT "Mouse Dll3: a novel divergent Delta gene which may complement the
function of other Delta homologues during early pattern formation in
the mouse embryo.";
RL Development 124:3065-3076(1997).
DR EMBL: Y11895; CAA72637.1; -
DR HSP; P00740; LIXA.
DR MGD; MGI:1096877; D113.
DR PFAM; PF00008; EGF; 6.
DR PRINTS; PR00010; EGFBL00D.
KW Signal; Glycoprotein.
FT SIGNAL 1 32 POTENTIAL.
SQ SEQUENCE 585 AA; 61129 MW; 0F5E6187 CRC32;

Query Match 12.8%; Score 74.5; DB 5; Length 1698;
Best Local Similarity 26.4%; Pred. No. 8.7;
Matches 23; Conservative 6; Mismatches 25; Indels 33; Gaps 5;

QY 68 ---AVTGCTC-----GSAC-GSWDVRAETTCCHCQAGMDWTGARC-----CRVQ 108
| : | : | : | : | : | : | : | : | : | : | : | : |
Db 305 LRCEVSVTCADGPCFNGGLCVGGEDPDSAYVCHCP---PGFQGSNCKRVRDCLSLQ 359

RESULT 15
ID O94438 PRELIMINARY; PRT; 1698 AA.
AC O94438;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE 185 KDA SILK PROTEIN.
GN SPI85.
OS Chironomus pallidivittatus (Midge).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Chironomidae; Chironomidae; Chironomidae; Chironomus.
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=SALIVARY GLAND;
RA CASE S.T., COX C., BELL W.C., HOFFMAN R.T., MARTIN J., HAMILTON R.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U54640; AAA99803.1; -
DR HSP; P18055; 2MRB.
DR PRINTS; PR00876; MINEMATODE.
SQ SEQUENCE 1698 AA; 186164 MW; 85E8E520 CRC32;

Query Match 12.8%; Score 74.5; DB 5; Length 1698;
Best Local Similarity 26.4%; Pred. No. 8.7;
Matches 23; Conservative 6; Mismatches 25; Indels 33; Gaps 5;
```


Oy 52 QSVTSRGLATCPRGFAVTGCT--CGSAC-----GSWDVRAETTCOC----- 94
I: | : | | | | | : | | : | | | | : | | | |
Db 1355 QTNWSOTCQCSCPCASGTCTGAQVWCCKACKVCYCPAQKKCDSPKTWD---ESSCQCPCPN 1411
Oy 94 -----AGMDWTGARC---CRVOP 108
I: | : | | | | | : | | : | | | | : | | | |
Db 1412 MRPPKGGCNAGRTWDDATCSEKCAAVP 1438

Search completed: May 26, 2000, 03:19:34
Job time: 10004 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2000, 00:06:17 ; Search time 39.3 Seconds
(without alignments)
65.092 Million cell updates/sec

Title: US-09-099-898-2

Perfect score: 581

Sequence: 1 MKALCLLLLPVGLLVSSKT.....CHCQCAGMDWTGRCRCRVQ 108

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match Length	DB ID	Description
1	581	100.0	108	1 W87710	A cysteine rich so
2	581	100.0	109	1 Y12933	Amino acid sequenc
3	318	54.7	114	1 W87707	A cysteine rich so
4	315	54.2	114	1 W87708	A cysteine rich so
5	284.5	49.0	111	1 W87709	A cysteine rich so
6	273.5	47.1	105	1 W87706	A cysteine rich so
7	245.5	42.3	111	1 W87705	A cysteine rich so
8	227	39.1	52	1 Y12613	Human 5' EST seque
9	202.5	34.9	111	1 W87704	A cysteine rich so
10	82	14.1	1872	1 W68510	Partial human Notc
11	82	14.1	2321	1 W49698	Human Notch3 prote
12	76.5	13.2	60	1 R9207	Nucleotide used in
13	73	12.6	131	1 W7444	Mouse tie receptor
14	72	12.4	1111	1 R91428	Kalinin/laminin 5
15	72	12.4	1193	1 R91427	Kalinin/laminin 5
16	71.5	12.3	519	1 R30826	The novel tyrosina
17	71.5	12.3	2476	1 W67738	Pig p105 zona pell
18	71	12.2	397	1 R27199	ALP of Lysobacter
19	69	11.9	1094	1 R39821	Truncated tie rece
20	68.5	11.8	2471	1 Y06816	Human Notch2 (hum
21	68	11.7	102	1 W24566	Serine protease C-
22	68	11.7	1122	1 R73954	Human tie tyrosine
23	68	11.7	1138	1 R39820	tie receptor kinas
24	67	11.5	915	1 Y13350	Amino acid sequenc
25	67	11.5	1801	1 W50895	Rat laminin B2 cha
26	66.5	11.4	488	1 R67757	Human fetal brain
27	66.5	11.4	524	1 R67758	Human fetal brain
28	66.5	11.4	670	1 R67759	Human fetal brain
29	66.5	11.4	769	1 R75352	Human fetal brain
30	66	11.4	177	1 R40167	Recombinant growth
31	66	11.4	297	1 W20066	Thielavia terrestr
32	65.5	11.3	1080	1 P50296	Mouse epidermal gr
33	65.5	11.3	1964	1 W95557	Mus musculus notch
34	65	11.2	289	1 R52633	Guinea pig PH-30,

35 65 11.2 1148 1 W87895 Human JAGGED2 prot
36 65 11.2 2707 1 W27161 Mouse receptor ME2
37 64.5 11.1 375 1 W4141 Thermostable alk1
38 64.5 11.1 2799 1 W81867 Human tumour suppr
39 64 11.0 197 1 W36951 Protein encoded by
40 64 11.0 213 1 R45359 Wheat germ aggluti
41 64 11.0 500 1 W94496 Human delta-2 prot
42 64 11.0 659 1 W80813 Nucleotide sequenc
43 64 11.0 685 1 W94507 Human delta-2 prot
44 64 11.0 685 1 W94507 Human delta-2 prot
45 64 11.0 688 1 P80361 Tick WGL+ antigen.

ALIGNMENTS

RESULT 1

W87710
ID W87710 standard; Protein; 108 AA.
AC W87710;
DT 09-MAR-1999 (first entry)
DE A cysteine rich soluble protein designated C23.
KW Cysteine rich soluble protein; CRSP; C23; cell development;
KW mammalian immune system; antibody; abnormal proliferation; cancer;
KW inflammation; degeneration; regeneration; atrophy.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Protein 1..18 /note= "mature protein"
PN W0989061-A1.
PD 23-DEC-1998.
PF 18-JUN-1998; U12236.
PR 03-OCT-1997; US-061641.
PR 19-JUN-1997; US-878730.
PR 19-JUN-1997; US-878878.
PA (SCHE) SCHERING CORP.
PI Franz-Bacon K, Gorman DM, McClanahan TK;
DR WPI: 99-095339/08.
DR N-PSDB; V84059.
PT New cysteine-rich soluble proteins - used to modulate proliferation,
PT differentiating, trafficking and development of cells, e.g. for
PT treating inflammation, cancer and degeneration
PS Claim 1: Page 18-19: 119pp; English.
CC The present sequence represents a cysteine rich soluble protein (CRSP)
CC designated C23. CRSP proteins, and their antagonists, are used to
CC modulate physiology, differentiation, trafficking and development
CC of cells (including those in culture), particularly cells of the
CC mammalian immune system. They are used for treatment of abnormal
CC proliferation (cancer, inflammation or degeneration), regeneration,
CC degeneration and atrophy. The proteins are also used to raise, or
CC detect, antibodies, to design oligonucleotides for library screening,
CC in drug screens and to isolate cognate receptors. The antibodies are
CC used for affinity purification of CRSP, to screen expression libraries,
CC to identify CRSP-expressing cells, as diagnostic immunoassay reagents,
CC to produce anti-idiotypic antibodies (useful for diagnosis), in
CC competitive drug screens, and as therapeutic modulators.
SQ Sequence 108 AA;

Query Match 100.0%; Score 581; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 8.3e-50;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKALCLLLLPVGLLVSSKTLCSMEERINERIQVAGSLIFRAISSIGLEQCVTSRGDL 60

Db 1 MKALCLLLLPVGLLVSSKTLCSMEERINERIQVAGSLIFRAISSIGLEQCVTSRGDL 60

OY 61 ATCPRGFAVTCGCGSAGSGWDVRAETTCQCQAGMDWTGRCRCRVQ 108

Db 61 ATCPRGFAVTCGCGSAGSGWDVRAETTCQCQAGMDWTGRCRCRVQ 108

RESULT 2

DR WPI: 99-095339/08.
 DR N-PSDB; V84057.
 PT New cysteine-rich soluble proteins - used to modulate proliferation,
 PT differentiation, trafficking and development of cells, e.g. for
 PT treating inflammation, cancer and degeneration
 PS Claim 1; Page 16; 119pp; English.
 CC The present sequence represents a cysteine rich soluble protein (CRSP)
 CC designated C18. CRSP proteins, and their (ant)agonists, are used to
 CC modulate physiology, differentiation, trafficking and development
 CC of cells (including those in culture), particularly cells of the
 CC mammalian immune system. They are used for treatment of abnormal
 CC proliferation (cancer, inflammation or degeneration), regeneration,
 CC degeneration and atrophy. The proteins are also used to raise, or
 CC detect, antibodies, to design oligonucleotides for library screening,
 CC in drug screens and to isolate cognate receptors. The antibodies are
 CC used for affinity purification of CRSP, to screen expression libraries,
 CC to identify CRSP-expressing cells, as diagnostic immunoassay reagents,
 CC to produce anti-idiotypic antibodies (useful for diagnosis), in
 CC competitive drug screens, and as therapeutic modulators.
 SQ Sequence 114 AA;

Query Match 54.2%; Score 315; DB 1; Length 114;
 Best Local Similarity 54.1%; Pred. No. 6.6e-24;
 Matches 60; Conservative 16; Mismatches 31; Indels 4; Gaps 1;

QY 1 MKALCLLL-...PVLGLLVSSKTLCSMEEAINEIRIOEVAGSLIFRAISSIGLECSQVTS 56
 DB 1 MKNLFLLLFLFVLGLLVSSKTLCSMEEAINEIRIOEVAGSLIFRAISSIGLECSQVTS 60

QY 57 RGLATCPRGFAVTCGTCGACGSDVRAETTCCHCCAGMDWTGACRCRV 107
 DB 61 RGLASCPGTTVTSCTGCGGSDVREDTCHCCGCSIDWTGACRCRV 111

RESULT 5
 W87709
 ID W87709 standard; Protein; 111 AA.
 AC W87709;
 DT 09-MAR-1999 (first entry)
 DE A cysteine rich soluble protein designated C10.
 KW Cysteine rich soluble protein; CRSP; C10; cell development;
 KW mammalian immune system; antibody; abnormal proliferation; cancer;
 KW inflammation; degeneration; regeneration; atrophy.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Protein 1..20
 FT /note= "mature protein"
 PN WO9858061-AL.
 PD 23-DEC-1998.
 PF 18-JUN-1998; U12236.
 PR 09-OCT-1997; US-061641.
 PR 19-JUN-1997; US-878730.
 PR 19-JUN-1997; US-878878.
 PA (SCHE) SCHERING CORP.
 PI Franz-Bacon K, Gorman DM, McClanahan TK;
 DR WPI: 99-095339/08.
 DR N-PSDB; V84058.
 PT New cysteine-rich soluble proteins - used to modulate proliferation,
 PT differentiation, trafficking and development of cells, e.g. for
 PT treating inflammation, cancer and degeneration
 PS Claim 1; Page 17; 119pp; English.
 CC The present sequence represents a cysteine rich soluble protein (CRSP)
 CC designated C10. CRSP proteins, and their (ant)agonists, are used to
 CC modulate physiology, differentiation, trafficking and development
 CC of cells (including those in culture), particularly cells of the
 CC mammalian immune system. They are used for treatment of abnormal
 CC proliferation (cancer, inflammation or degeneration), regeneration,
 CC degeneration and atrophy. The proteins are also used to raise, or
 CC detect, antibodies, to design oligonucleotides for library screening,
 CC in drug screens and to isolate cognate receptors. The antibodies are
 CC used for affinity purification of CRSP, to screen expression libraries,
 CC to identify CRSP-expressing cells, as diagnostic immunoassay reagents,
 CC to produce anti-idiotypic antibodies (useful for diagnosis), in
 CC competitive drug screens, and as therapeutic modulators.

CC to produce anti-idiotypic antibodies (useful for diagnosis), in
 CC competitive drug screens, and as therapeutic modulators.
 SQ Sequence 111 AA;

Query Match 49.0%; Score 284.5; DB 1; Length 111;
 Best Local Similarity 49.0%; Pred. No. 5.9e-21;
 Matches 51; Conservative 21; Mismatches 31; Indels 1; Gaps 1;

QY 4 LCLLLPVLGLLVSSKTLCSMEEAINEIRIOEVAGSLIFRAISSIGLECSQVTSRGDLAT 62
 DB 7 LLLILPVLGLLVSSKTLCSMEEAINEIRIOEVAGSLIFRAISSIGLECSQVTSRGDLAT 66

QY 63 CPRGFVAVTCGTCGACGSDVRAETTCCHCCAGMDWTGACRCRV 106
 DB 67 CPAGMVTGACGCGGSDVQLTTCCHCCGSDVVDWTATACCHL 110

RESULT 6
 W87706
 ID W87706 standard; Protein; 105 AA.
 AC W87706;
 DT 09-MAR-1999 (first entry)
 DE A cysteine rich soluble protein designated C18.
 KW Cysteine rich soluble protein; CRSP; C18; cell development;
 KW mammalian immune system; antibody; abnormal proliferation; cancer;
 KW inflammation; degeneration; regeneration; atrophy.
 OS Mus sp.
 FH Key Location/Qualifiers
 FT Protein 1..19
 FT /note= "mature protein"
 PN WO9858061-AL.
 PD 23-DEC-1998.
 PF 18-JUN-1998; U12236.
 PR 09-OCT-1997; US-061641.
 PR 19-JUN-1997; US-878730.
 PR 19-JUN-1997; US-878878.
 PA (SCHE) SCHERING CORP.
 PI Franz-Bacon K, Gorman DM, McClanahan TK;
 DR WPI: 99-095339/08.
 DR N-PSDB; V84055.
 PT New cysteine-rich soluble proteins - used to modulate proliferation,
 PT differentiation, trafficking and development of cells, e.g. for
 PT treating inflammation, cancer and degeneration
 PS Claim 1; Page 14; 119pp; English.
 CC The present sequence represents a cysteine rich soluble protein (CRSP)
 CC designated C18. CRSP proteins, and their (ant)agonists, are used to
 CC modulate physiology, differentiation, trafficking and development
 CC of cells (including those in culture), particularly cells of the
 CC mammalian immune system. They are used for treatment of abnormal
 CC proliferation (cancer, inflammation or degeneration), regeneration,
 CC degeneration and atrophy. The proteins are also used to raise, or
 CC detect, antibodies, to design oligonucleotides for library screening,
 CC in drug screens and to isolate cognate receptors. The antibodies are
 CC used for affinity purification of CRSP, to screen expression libraries,
 CC to identify CRSP-expressing cells, as diagnostic immunoassay reagents,
 CC to produce anti-idiotypic antibodies (useful for diagnosis), in
 CC competitive drug screens, and as therapeutic modulators.
 SQ Sequence 105 AA;

Query Match 47.1%; Score 273.5; DB 1; Length 105;
 Best Local Similarity 48.8%; Pred. No. 6.5e-20;
 Matches 51; Conservative 12; Mismatches 35; Indels 7; Gaps 2;

QY 4 LCLLLPVLGLLVSSKTLCSMEEAINEIRIOEVAGSLIFRAISSIGLECSQVTSRGDLA 61
 DB 5 LCLFLVLVFLPLVPGNAQCSFSLVDQIKKALSRQEPKTIIS-----CISVTSSGRLA 59

QY 62 TCRPGFAVTCGTCGACGSDVRAETTCCHCCAGMDWTGACRCRV 106
 DB 60 SCAGMVTGACGCGGSDVIRNGTCHCCGSDVVDWTATACCHL 104

```

RESULT 7
W87705 ID W87705 standard; Protein; 111 AA.
AC W87705; 1999 (first entry)
DE A cysteine rich soluble protein designated C2b.
KW Cysteine rich soluble protein; CRSP; C2b; cell development;
KW mammalian immune system; antibody; abnormal proliferation; cancer;
KW inflammation; degeneration; regeneration; atrophy.
OS Mus sp.
FH Key Location/Qualifiers
FT Protein 1..23 /note= "mature protein"
FT WO9858061-A1.
PN 23-DEC-1998.
PD 18-JUN-1998; UI2236.
PF 09-OCT-1997; US-061641.
PR 19-JUN-1997; US-878730.
PR 19-JUN-1997; US-878878.
PA (SCHE ) SCHERING CORP.
PI Franz-Bacon K, Gorman DM, McLanahan TK;
DR N-PSDB; V84054.
DR New cysteine-rich soluble proteins - used to modulate proliferation,
PT differentiation, trafficking and development of cells, e.g. for
PT treating inflammation, cancer and degeneration
PS Claim 1; Page 13; 119pp; English.
CC The present sequence represents a cysteine rich soluble protein (CRSP)
CC designated C2b. CRSP proteins, and their (ant)agonists, are used to
CC modulate physiology, differentiation, trafficking and development
CC of cells (including those in culture), particularly cells of the
CC mammalian immune system. They are used for treatment of abnormal
CC proliferation (cancer, inflammation or degeneration), regeneration,
CC degeneration and atrophy. The proteins are also used to raise, or
CC detect, antibodies, to design oligonucleotides for library screening,
CC in drug screens and to isolate cognate receptors. The antibodies are
CC used for affinity purification of CRSP, to screen expression libraries,
CC to identify CRSP-expressing cells, as diagnostic immunoassay reagents,
CC to produce anti-idiotypic antibodies (useful for diagnosis), in
CC competitive drug screens, and as therapeutic modulators.
SQ Sequence 111 AA;

Query Match 42.38; Score 245.5; DB 1; Length 111;
Best Local Similarity 40.28; Pred. No. 3.7e-17;
Matches 43; Conservative 21; Mismatches 32; Indels 11; Gaps 2;

QY 6 LLLPLVGLVSSKTLCSMEERINERIOEVAG-----SLIFRAISSIGLCQSVTSRGD 59
DB 9 LICISLQLMPVNVTEGTLESIVKVKELLANRDCCPSTVTKTFS-----CITSASGR 63

QY 60 LATCPRGFVNTGCTCGSAGSDVRAETTHCCQAGMDWTGACRCV 106
DB 64 LASCPSGMTVTGACGCGGSDIRDGNTHCQCSTMDWATARCQQL 110

RESULT 8
Y12613 ID Y12613 standard; Protein; 52 AA.
AC Y12613;
DE Human 5' EST secreted protein SEQ ID NO: 278 from WO 9006553.
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; antiinflammatory; tumour inhibition; antitumour.
OS Homo sapiens.
PN WO9906553-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IB1237.

```

```

PR 01-AUG-1997; US-905051.
PA (GEST ) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
DR WPI; 99-153783/13.
DR N-PSDB; X41471.
DR New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries derived from umbilical cord, lymph ganglia,
PT lymphocytes and placental tissue
PS Claim 34; Page 375; 41pp; English.
CC X41379 to X41526 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y12521 to Y12668,
CC respectively. The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, haematopoiesis regulating activity, tissue growth regulating
CC activity, reproductive hormone regulating activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, antiinflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC the signal peptide can be used for directing extracellular secretion of
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell.
SQ Sequence 52 AA;

Query Match 39.18; Score 227; DB 1; Length 52;
Best Local Similarity 100.0%; Pred. No. 1e-15;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKALCLLLPVGLVSSKTLCSMEERINERIOEVAGSLIFRAISSIG 48
DB 1 MKALCLLLPVGLVSSKTLCSMEERINERIOEVAGSLIFRAISSIG 48

RESULT 9
W87704 ID W87704 standard; Protein; 111 AA.
AC W87704;
DT 09-MAR-1999 (first entry)
DE A cysteine rich soluble protein designated C2.
KW Cysteine rich soluble protein; CRSP; C2; cell development;
KW mammalian immune system; antibody; abnormal proliferation; cancer;
KW inflammation; degeneration; regeneration; atrophy.
OS Mus sp.
FH Key Location/Qualifiers
FT Protein 1..23 /note= "mature protein"
FT WO9858061-A1.
PN 23-DEC-1998.
PD 18-JUN-1998; UI2236.
PF 09-OCT-1997; US-061641.
PR 19-JUN-1997; US-878730.
PR 19-JUN-1997; US-878878.
PA (SCHE ) SCHERING CORP.
PI Franz-Bacon K, Gorman DM, McLanahan TK;
DR WPI; 99-095339/08.
DR N-PSDB; V84053.
DR New cysteine-rich soluble proteins - used to modulate proliferation,
PT differentiation, trafficking and development of cells, e.g. for
PT treating inflammation, cancer and degeneration
PS Claim 1; Page 12-13; 119pp; English.
CC The present sequence represents a cysteine rich soluble protein (CRSP)
CC designated C2. CRSP proteins, and their (ant)agonists, are used to
CC modulate physiology, differentiation, trafficking and development
CC of cells (including those in culture), particularly cells of the
CC mammalian immune system. They are used for treatment of abnormal
CC proliferation (cancer, inflammation or degeneration), regeneration,
CC degeneration and atrophy. The proteins are also used to raise, or
CC detect, antibodies, to design oligonucleotides for library screening,

```

CC developmental cascades of neurogenic genes. Mutated Notch3 proteins
CC are thought to be involved in neurological disorders, especially of the
CC cerebral autosomal dominant arteriopathy with subcortical infarcts and
CC leukoencephalopathy (CADASIL) type. Blocking expression of a mutated
CC Notch3 gene or by substitution therapy with non-mutated Notch3 gene or
CC protein can be used to treat CADASIL or related disorders.
CC Sequence 1872 AA;
SQ

```

Query Match      14.1%; Score 82; DB 1; Length 1872;
Best Local Similarity 29.9%; Pred. No. 6.6;
Matches 23; Conservative 7; Mismatches 23; Indels 24; Gaps

QY 47 IGLEQSVTSRGDLATCPRGFAVTGCTCGSAGSGWDVRAETTC----- 90
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1234 VGVPCCQ-TPRGPRCACPPG--LSGPGSCRSPGPGGASNAACAAACLHGGSCRPA 1290
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 90 ---HCQCAGMDWTGARC 103
      _ _ _ ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1291 PFFRCACA-QGWTGPRC 1306
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

[illegible]

30-DEC-1996 (first entry)
DE Nucleotide used in production of MSH/MoMuLV chimeric sequence.
DE Moloney murine leukaemia virus; gp70; 4070A retrovirus; retrovirus;
KW 10A1 murine leukaemia virus; NZB-9-1 murine leukaemia virus;
KW polyporphic MX27 provirus; targetted drug delivery; gene therapy;
KW single chain antibody; envelope protein; ss.
OS Synthetic.
PN WO9630504-A1.
PD 03-OCT-1996.
PF 22-MAR-1996; U03908.
PR 24-MAR-1995; US-409648.
PA (GENE-) GENETIC THERAPY INC.
PA (UYSC-) UNIV SOUTHERN CALIFORNIA.
PI Anderson W, Chiang YL, Januszski M, Mackrell AJ;
PI Zhao Y;
DR WPI; 96-455352/45.
PT Cell-targetted retroviral vector particles - having envelope protein
PT modified with targeting polypeptide
PS Example 2; Page 36; 73pp; English.
CC Cell targetted retroviral vector particles can be used in gene
CC therapy to deliver a heterologous gene to a target cell for
CC expression of a heterologous polypeptide in that cell. The cell
CC targetted retroviral vector particles comprise an envelope protein
CC which is modified to contain a targeting polypeptide (a single chain
CC antibody), or in the case of moloney murine leukaemia virus
CC (MoMuLV), alpha melanotropin-stimulating hormone (MSH). Two
CC oligonucleotides (R98207, R98208) were used to substitute sequences in
CC MoMuLV for MSH sequences. This oligonucleotide was used to replace
CC residues 574-591 of MoMuLV envelope protein (See W04248).
SQ Sequence 60 AA;

Query Match 13.2%; Score 76.5; DB 1; Length 60;
Best Local Similarity 35.1%; Pred. No. 0.51;
Matches 20; Conservative 3; Mismatches 19; Indels 15; Gaps 3;

[illegible]

RESULT 13

W74444 standard; Protein; 131 AA.

ID W74444

AC W74444;

DT 11-MAY-1999 (first entry)

DE Mouse Tie receptor tyrosine kinase fragment.

KW Tie gene; receptor tyrosine kinase; promoter; gene expression;

KW human gene therapy; growth factor.

OS Mus sp.

PN US5877020-A.

PD 02-MAR-1999.

PF 31-MAY-1996; 650598.

PR 31-MAY-1996; US-650598.

PR 22-SEP-1994; US-310717.

PA (UYHE-) UNIV HELSINKI LICENSING LTD.

PI Alitalo K;

DR WPI; 99-189653/16.

PT Tie receptor tyrosine kinase promoter - for directing expression of

PT recombinant DNA in endothelial cells

PS Example 1; Column 6; 26pp; English.

CC This sequence represents a fragment of the mouse Tie receptor tyrosine

CC kinase. The invention relates to Tie receptor expression of recombinant DNA

CC The promoters are useful for directing expression of recombinant DNA

CC sequences in endothelial cells. The promoters are useful for production

CC of proteins and peptides which act as anticoagulants, vasodilator

CC inhibitors of thrombosis or restenosis into endothelial cells, blood and

CC tissues. The promoters are useful for directing expression of proteins

CC and peptides for human gene therapy, antigens and markers for endothelial

CC cell tagging, and antisense RNA constructs for use in endothelial cells

CC in vivo and in vitro. The promoters, and vectors and host cells

CC containing them, are useful in gene therapy for promoting expression of

CC various growth factors of receptors or their domains.

```

SQ      Sequence      131 AA;

Query Match      12.6%; Score 73; DB 1; Length 131;
Best Local Similarity 28.8%; Pred. No. 2.6;
Matches 23; Conservative 4; Mismatches 19; Indels 34; Gaps 5;

QY      59  DLATCPRGFAVTGCTCGSAC-----GSDVRAETTC-----HCQCAGMDWT 99
          |  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db       19  DGCVCPPGF--TGTRCQACREGRFGSCQFCPGTAGCRGLTFCLPDPYGCGR 75
          |  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||

QY      100 GARC-----CRVQ 107
          ||::|
Db       76  GSQCQECACPDHFGADCRLQ 95
          ||::|

```

RESULT 14

R91428 R91428 standard; Protein; llll AA.

ID ID R91428; AC R91428;

DT DT 13-NOV-1996 (first entry)

DE DE Kallinin/laminin 5 gamma-2 chain (alternative form).

KW Kallinin; laminin; laminin 5 gamma-2 chain (alternative form);

KW Kallinin; laminin; epidermolysis bullosa; junctional; probe;

KW detection; inhibit; monitor; malignancy.

OS Homo sapiens.

PN WO9610646-A1.

PD 11-APR-1996.

PF 04-OCT-1995; E03918.

PR PR 04-OCT-1994; US-317450.

PA (TRYG/) TRYGGVASON K.

PI Kallunki P. Pyke C. Tryggvason K;

DR WFI; 96-209366/21.

DR N-PSDB; TL3323.

DR Detection of kallinin or laminin 5 expression in cells - useful to

PT detect, monitor and inhibit the invasive growth of cell in tissue,

PT partic. malignant tissue

PS Disclosure; Fig 4B; 37pp; English.

CC The present sequence is an alternative form of kallinin/laminin 5 gamma-2

CC chain (see R91427). The gamma-2 chain is of importance to patients

CC suffering from epidermolysis bullosa, esp. the junctional form (JEB).

CC Probes and antisense gamma-2 sequences derived from this sequence can be

CC used to detect, monitor and inhibit the invasive growth of cells in

CC tissue, partic. malignant tissue.

SQ Sequence llll AA;

RESULT	15
R91427	
ID	R91427 standard; Protein; 1193 AA.
AC	R91427;
DT	13-NOV-1996 (first entry)
DE	Kalinin/laminin 5 gamma-2 Chain.
KW	kalinin; laminin; epidermolysis bullosa; junctional; probe;
KW	detection; inhibit; monitor; malignancy.
OS	Homo sapiens.
PN	W09610646-A1.

```

PD 11-APR-1996.
PF 04-OCT-1995; E03918.
PR 04-OCT-1994; US-317450.
PA (TRYG/) TRYGGVASON K.
PI Kallunki P, Pyke C, Tryggvason K;
DR WPI; 96-209366/21.
DR N-PSDB; T13323.
PT Detection of kalinin or laminin 5 expression in cells - useful to
PT detect, monitor and inhibit the invasive growth of cell in tissue,
PT partic. malignant tissue
PS Disclosure; Fig 4A; 37pp; English.
CC The present sequence is the kalinin/laminin 5 gamma-2 chain. The gamma-2
CC chain is of importance to patients suffering from epidermolysis bullosa,
CC esp. the junctional form (JEB). Probes and antisense gamma-2 sequences
CC derived from this sequence can be used to detect, monitor and inhibit
CC the invasive growth of cells in tissue, partic. malignant tissue.
SQ Sequence 1193 AA;

Query Match 12.4%; Score 72; DB 1; Length 1193;
Best Local Similarity 25.4%; Pred. No. 38;
Matches 31; Conservative 14; Mismatches 47; Indels 30; Gaps 8;

QY 4 LCLLLPLVLGLVSKTLC-----SMEEAINEIRIQEVAGSLIFRAI-----SSIGLEQSV 54
Db 10 LCFSLLLPAARATSRREVCDKNGKSRQCFDREHLHROTGN-GFCLNCNDNTDGIHCEK- 68

QY 55 TSGDLATCPRGF-----AVTGCTGSGAGWDVRAETTCQC-AGMDWTGACRCV 106
Db 68 -----CKNGFYRHRERDRCLPCNCSK-GSLSARCDNSGRCSKPGV--TGARCDRC 116

QY 107 QP 108
Db 117 LP 118

Search completed: May 26, 2000, 03:02:14
Job time: 10557 sec

```


Result No.	Query Match	Score	Length	DB	ID	Description
1	453	100.0	453	1	V84059	CDNA encoding a cy
2	420.6	92.8	553	1	X51720	DNA encoding a hum
3	194	42.8	229	1	X14171	Human secreted pro
4	158.4	35.0	572	1	V84057	CDNA encoding a cy
5	152.4	33.6	560	1	V84056	CDNA encoding a cy
6	92.8	20.5	603	1	V84058	CDNA encoding a cy
7	78	17.2	554	1	V84055	CDNA encoding a cy
8	64.4	14.2	574	1	V84054	CDNA encoding a cy
9	54.4	12.0	527	1	V84053	CDNA encoding a cy
10	40.6	9.0	985	1	V44439	Mycobacterium tube
11	40.6	9.0	985	1	V64548	M. tuberculosis im
12	39	8.6	8438	1	Q73500	DNA encoding Pseud
13	38.6	8.5	114955	1	X53491	Human adenosine A1
14	36.6	8.1	3415	1	T36481	Human integrin bet
15	36.6	8.1	11820	1	V18130	Human chromosome 1
16	35.2	8.0	2932	1	Q25388	TXA2 receptor gene
17	35.2	7.8	329	1	V44425	Mycobacterium tube
18	35.2	7.8	329	1	V84534	M. tuberculosis im
19	34.8	7.7	30001	1	T61016	Total DNA sequenc
20	34.8	7.7	30001	1	X05110	S. aureofaciens DN
21	34.8	7.7	114955	1	X53491	Human adenosine A1
22	34.8	7.7	117213	1	V82176	HSV-2 strain SB5 C
23	34.4	7.6	201	1	N70195	Streptomyces prote
24	34.2	7.5	1035	1	T49316	Hypersensitive res
25	34.2	7.5	1035	1	V36430	Hypersensitive res
26	34.2	7.5	1035	1	V39975	Pseudomonas solana
27	34.2	7.5	1035	1	V34609	Pseudomonas solana
28	34.2	7.5	1035	1	V83991	DNA encoding a hyp
29	34	7.5	201	1	X19063	Human PPAR-gamma-1
30	34	7.5	384	1	V23147	Rat Puralpha like
31	34	7.5	1545	1	V41906	Nucleotide sequenc
32	34	7.5	1687	1	T39798	Human clone 56 gen
33	33.8	7.5	535	1	V44428	Mycobacterium tube
34	33.8	7.5	535	1	V64533	M. tuberculosis im


```

RESULT      4
V84057
ID          V84057 standard; cDNA; 572 BP.
AC          V84057;
DT          09-MAR-1999 (first entry)
DE          cDNA encoding a cysteine rich soluble protein designated C19.
KW          Cysteine rich soluble protein; CRSP; C19; cell development; cancer;
KW          mammalian immune system; antibody; abnormal proliferation; cancer;
KW          inflammation; degeneration; regeneration; atrophy; ss.
OS          Rattus sp.
Locus       Location/Qualifiers
Key         29..373
CD5         /tag= a
FT          FT
FT          /product= C19
FT          29..79
FT          /tag= b
FT          80..370
FT          /tag= c
FT          159..160
FT          /tag= d
FT          /note= "Intron present between these nucleotides in
FT          the genomic DNA"
FT          236..237
FT          /tag= e
FT          /note= "intron present between these nucleotides in
FT          the genomic DNA"
FT          WO9858061-A1.
FN          23-DEC-1998.
PD          18-JUN-1998; U12236.
PE          09-OCT-1997; US-061641.
PR          19-JUN-1997; US-878730.
PR          19-JUN-1997; US-878878.
PA          (SCHE ) SCHERING CORP.
PI          Franz-Bacon K, Gorman DM, McClanahan TK;
DR          WFI; 99-095339/08.
DR          P-PSDB; W87708.

```

```

RESULT      5
V84056      V84056 standard; cDNA; 560 BP.
ID          V84056;
AC          09-MAR-1999 (first entry)
DT          DE cDNA encoding a cysteine rich soluble protein designated C19.
KW          Cysteine rich soluble protein; CRSP; C19; cell development;
KW          mammalian immune system; antibody; abnormal proliferation; cancer;
KW          inflammation; degeneration; regeneration; degeneration; atrophy; ss.
KW          Mus sp.
OS          Location/Qualifiers
Key         64. .408
CDS         /*tag= a
            /product= C19
FT          64. .123
FT          /*tag= b
FT          124. .405
FT          /*tag= c
FT          193. .194
FT          /*tag= d
FT          /note= "Intron present between these nucleotides in
FT          the genomic DNA"
FT          271. .272
FT          /*tag= e
FT          misc_feature
FT          misc_feature

```


Db	285	TCCTGCTGGATGGTTGCTCACTGGATGCTCTGTGGCTATGCTGTGGATCTGGATCTGGGATAT	344
QY	295	GCCTGCCGAGACACATGTCACTGCCAGTGCCTGGGCGCATGGATGGACCGGACGGCGTGTG	354
Db	345	CCGGAATGGAATACTTGCACCTGCCAGTGCCTCACTCATGGACTGGGCGCTCTCTCCGCGTGTG	404
QY	355	CTGCTGCTGTG	364
Db	405	CTGCCGAATG	414
<p>RESULT 8</p> <p>V84054</p> <p>ID V84054 standard; cDNA; 574 BP.</p> <p>AC V84054;</p> <p>DT 09-MAR-1999 (first entry)</p> <p>DE cDNA encoding a cysteine rich soluble protein designated C2b.</p> <p>KW Cysteine rich soluble protein; CRSP; C2b; cell development;</p> <p>KW mammalian immune system; antibody; abnormal proliferation; cancer;</p> <p>OS inflammation; degeneration; regeneration; atrophy; ss.</p> <p>OS Mus sp.</p> <p>Key</p> <p>FT CDS</p> <p>FT 70..405</p> <p>FT /*tag= a</p> <p>FT /product= C2b</p> <p>FT sig_peptide</p> <p>FT 70..138</p> <p>FT /*tag= b</p> <p>FT mat_peptide</p> <p>FT 139..402</p> <p>FT /*tag= c</p> <p>FT misc_feature</p> <p>FT 196..197</p> <p>FT /*tag= d</p> <p>FT /note= "intron present in genomic DNA"</p> <p>FT 277..278</p> <p>FT /*tag= e</p> <p>FT /note= "intron present in genomic DNA"</p> <p>FT</p> <p>PN W09858061-Al.</p> <p>PD 23-DEC-1998.</p> <p>PF 18-JUN-1998; U12236.</p> <p>PR 09-OCT-1997; US-061641.</p> <p>PR 19-JUN-1997; US-878730.</p> <p>PR 19-JUN-1997; US-878878.</p> <p>PA (SCHE) SCHERING CORP.</p> <p>PA Franz-Bacon K, Gorman DM, McClanahan TK;</p> <p>PI WPI: 99-095339/08.</p> <p>DR P-PSDB; W87705.</p> <p>DR New cysteine-rich soluble proteins - used to modulate proliferation,</p> <p>PT differentiation, trafficking and development of cells, e.g. for</p> <p>PT treating inflammation, cancer and degeneration</p> <p>PT Claim 16; Page 13; 119pp; English.</p> <p>PS The present sequence encodes a cysteine rich soluble protein (CRSP)</p> <p>CC designated C2b. CRSP proteins, and their (ant)agonists, are used to</p> <p>CC modulate physiology, differentiation, trafficking and development</p> <p>CC of cells (including those in culture), particularly cells of the</p> <p>CC mammalian immune system. They are used for treatment of abnormal</p> <p>CC proliferation (cancer, inflammation or degeneration), regeneration,</p> <p>CC degeneration and atrophy. The proteins are also used to raise, or</p> <p>CC detect, antibodies, to design oligonucleotides for library screening,</p> <p>CC in drug screens and to isolate cognate receptors. The antibodies are</p> <p>CC used for affinity purification of CRSP, to screen expression libraries</p> <p>CC to identify CRSP-expressing cells, as diagnostic immunoassay reagents</p> <p>CC to produce anti-idiotypic antibodies (useful for diagnosis), in</p> <p>CC competitive drug screens, and as therapeutic modulators.</p> <p>CC Sequence 574 BP; 161 A; 139 C; 122 G; 152 T;</p>			
<p>Query Match 14.2%; Score 64.4; DB 1; Length 574;</p> <p>Best Local Similarity 62.3%; Pred. No. 6.5e-08;</p> <p>Matches 101; Conservative 0; Mismatches 61; Indels 0; Gaps</p>			
QY	203	AGCGTCACCTCCAGGGGACCTGGCTACTTGCCTCCCGAGGCTTCGCCCTCACCGGCTGC	262
Db	238	AGTACCGGCTTCAGCGACAGTGGCTCCTGTCTTGGATGACTGTACTGTTGT	297


```
AC V64548;
DT 27-JAN-1999 (first entry)
DE M. tuberculosis immunogenic polypeptide XP25 3'-end DNA.
KW tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW vaccine; pharmaceutical; infection; diagnosis; ss.
OS Mycobacterium tuberculosis.
PN W09816646-A2.
PD 23-APR-1998.
PE 07-OCT-1997; U18293.
PR 13-MAR-1997; US-818112.
PR 11-OCT-1996; US-730510.
PA (CORI-) CORIXA CORP.
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ,
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
DR WPI; 98-261042/23
DR Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and for diagnosis, treatment and prevention of tuberculosis
PS Claim 4; Page 176; 230pp; English.
CC This sequence encodes an immunogenic portion of a soluble Mycobacterium
CC tuberculosis (MT) antigen which can be used in a method for inducing
CC protective immunity against tuberculosis (TB). This sequence can be
CC formulated into vaccines and/or pharmaceutical compositions for
CC immunising against M. tuberculosis infection or may be used for the
CC diagnosis of tuberculosis.
CC Sequence 985 BP; 138 A; 330 C; 423 G; 94 T;
SQ

Query Match 9.0%; Score 40.6; DB 1; Length 985;
Best Local Similarity 45.3%; Pred. No. 0.079;
Matches 148; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

QY 111 GCTCCATGGAAGACCATCAATGAGAGGATCCAGAGGTGCGCGGTCCCTATATTTA 170
-D 111 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
98 GCGCCGCGGTGACGGCGGCAAGGCGGACCCGAGGTGCGCGGTGCGCGCGGACA 157
QY 171 GGGCAATACAGCATGTGGCTGTGAGTGCCAGAGCGTCACCTCCAGGGGGACCTGGCTA 230
-D 171 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
158 ACCCCACCGGTGTGTTTGCCTGCGGTGCGCGGCGGCGACAGGTGCGCGCGCGCG 217
QY 231 CTGCCCCCGAGGCTTCGCGTCCACCGGTGCACTGTGCTCGCGCTGTGCTGCTGGG 290
-D 231 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
218 CGCGGGCGCGCGGCGGACCGGTACCGGCGGCGGCGGCGGCGGCGGCGGCGGTA 277
QY 291 ATGTGCGCGCGGAGACACATGTCATGTCAGTCCGCGGGGATGAGTACCGCGGCGG 350
-D 291 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
278 GTGACGATCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGTC 337
QY 351 GCTGCTGCTGTGACGCTGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 410
-D 351 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
338 TCGGCTGCGGCTCTCGCGGCTTTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGA 397
QY 411 GCTCCAGTCCGGAGGGTTTCGCGGGG 437
-D 411 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
398 GCGCCGCGCGCGGCGGCGATCAACGGGG 424

RESULT 12
Q73500
ID Q73500 standard; DNA; 8438 BP.
AC Q73500;
DT 15-MAY-1995 (first entry)
DE DNA encoding Pseudorabies virus large latency transcript.
KW Pseudorabies virus; PRV; LLV; large latency transcript;
KW attenuated virus; vaccine; early protein 0; EP0; HSV-1 ICPO;
KW protecting animals; deletion mutants; swine; ds.
OS Pseudorabies virus.
FS Key Location/Qualifiers
FT misc_feature 1..7013
FT /tag= a
FT /note= "derived from PRV strain InFh"
FT misc_feature 7014..8425
FT /tag= b
```

```
FT /note= "derived from PRV strain Ka"
FT 622..6498
FT /tag= c
FT /note= "encodes predicted amino acid sequence of ORF2"
FT 1..6
FT /tag= d
FT 34
FT /tag= e
FT /note= "RNA cap site"
FT 8382..8387
FT /tag= f
FT
FN US5352596-A.
PD 04-OCT-1994.
PR 11-SEP-1992; 945283.
PR 11-SEP-1992; US-945283.
PA (USDA ) US SEC OF AGRIC.
PI Cheung AK, Wesley RD;
DR WPI; 94-316187/39.
DR P-PSDB; R60620.
PT New pseudorabies virus mutants for use in vaccine - having a
PT deletion and/or insertion in the early protein O gene or large
PT latency transcript gene.
PS Disclosure; Column 15-30; 43pp; English.
CC Q73500 shows the Pseudorabies virus (PRV) large latency transcript
CC (LLT). The basic sequence is derived from PRV strain InFh and PRV
CC strain Ka. The LLT overlaps and is transcribed in the opposite
CC orientation with respect to the EP0 (early polypeptide 0) and the
CC immediately early gene (IE180). EP0 is nonessential for replicatio,
CC LLT is the only gene expressed during PRV latency, and the IE180
CC gene is absolutely necessary for PRV replication. However there are
CC 2 copies of IE180 in the genome. It is expected that PRV lacking one
CC of the IE180 copies is viable. Deletions in the non-overlapping
CC regions of these 3 genes will generate single deletion mutants.
CC while deletions in overlapping regions will generate double deletion
CC mutants. The invention is concerned with the construction of attenuated
CC viruses which have a reduced ability to reactivate from latency. This
CC can be achieved by functionally disabling the expression of the EP0
CC gene, or by disrupting the synthesis of the LLT, or both. (See also
CC Q73501 and R60620-24)
SQ Sequence 8438 BP; 1141 A; 2916 C; 3327 G; 1054 T;

Query Match 8.6%; Score 39; DB 1; Length 8438;
Best Local Similarity 51.6%; Pred. No. 0.31;
Matches 114; Conservative 0; Mismatches 105; Indels 2; Gaps 1;

QY 217 GGGGACACTGCTACTTGCCTCCGAGGCTTCCTGTCACCGGTGCTGCTGCTGCTGCT 276
-D 217 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5479 GCGGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 5538
QY 277 CTGTGGCTCTGCGGATGTGCGGCGCGAGACACATGCTACTGCCAGTGGCGGCGATGGA 336
-D 277 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5539 GCGCGGCGGCGAGTGGGCGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5598
QY 337 CTGGACC--GGAGCGCGTGTGCTGTGTCAGCCCTGAGGTCGCGCGAGCGCGTGCAC 394
-D 337 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5599 CGAGACAGGAGGACGAGGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5658
QY 395 AGCGCGGCGGAGGCGGCTCCAGGTCCGAGGGGTTGCGGG 435
-D 395 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5659 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5699
```

```
RESULT 13
X53491
ID X53491 standard; DNA; 114955 BP.
AC X53491;
DT 05-JUL-1999 (first entry)
DE Human adenosine A1 receptor antisense oligonucleotide fragment.
KW Antisense oligonucleotide; multiple target; antisense treatment;
KW impaired respiration; inflammation; lung disease;
KW pulmonary vasoconstriction; inflammation; allergic rhinitis;
KW acute asthma; allergy; asthma; impeded respiration;
```


KW respiratory distress syndrome; pain; cystic fibrosis;
 KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;
 KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
 KW colon cancer; breast cancer; lung cancer; pancreatic cancer;
 KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
 KW prostate cancer; ss.
 OS Synthetic.
 PN W09913886-AL.
 PD 25-MAR-1999.
 PF 17-SEP-1998; U19419.
 PR 09-JUN-1998; US-093972.
 PR 17-SEP-1997; US-059160.
 PA (UYEC-) UNIV EAST CAROLINA.
 PI NYCE JW;
 DR WPI: 99-229400/19.
 PT New antisense oligonucleotides used in treatment of, e.g. pulmonary
 PT vasoconstriction
 PS Disclosure: Page 37; 120pp; English.
 CC The specification describes antisense oligonucleotides (X52869-X55271)
 CC directed against at least 2 mRNAs selected from target genes, coding and
 CC non-coding regions of RNAs corresponding to target genes, gene
 CC initiation codons, genomic flanking regions, intron-exon borders, the
 CC 5'-end, the 3'-end and the juxta-section between coding and non-coding
 CC regions and all segments of RNAs encoding proteins associated with one
 CC or more diseases, conditions or mixtures. The antisense oligonucleotides
 CC may be derived from sequences X55272-74. These multiple target
 CC oligonucleotides (specifically X55180-271) can be used for the antisense
 CC treatment of diseases and conditions. Typical diseases and conditions
 CC are those associated with impaired respiration and inflammation,
 CC including lung diseases, pulmonary vasoconstriction, inflammation,
 CC allergic rhinitis, acute asthma, allergies, asthma, impeded respiration,
 CC respiratory distress syndrome, pain, cystic fibrosis, pulmonary
 CC hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
 CC carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic
 CC cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic
 CC metastases, as well as all types of cancers which may metastasize or have
 CC metastasized to the lungs, including breast and prostate cancer.
 CC Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T;
 SQ Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T;

Query Match 8.5%; Score 38.6; DB 1; Length 114955;
 Best Local Similarity 36.0%; Pred. No. 0.65;
 Matches 93; Conservative 19; Mismatches 146; Indels 0; Gaps 0;
 QY 180 GCAGCATTGGCTGAGTCCAGAGCGTCACTCCAGGGGGACCTGGCTACTTGGCCCC 239
 Db 103818 GCGGGCTCGGGCGGCGGCSNNNDNCCGTCTGTGGCGCGTGGCGGCGGCSNNNDNCCGT 103877
 QY 240 GAGGCTTCGCGGTCACCGGCTGCACCTGTGGCTCCGCTGTGGATGTGGCG 299
 Db 103878 BTGGCGGCTCGGGCGGCSNNNDNCCGTCTGTGGCGCGTGGCGGCGGCSNNNDNCCGT 103937
 QY 300 CCGAGACCATGTCTACTCCAGTGGCGGCGATGGACTGGACCGGCGCGCTGCTGTC 359
 Db 103938 TGGCGGCGTGGCGGCGGCSNNNDNCCBTGGCGGCTGGCGGCGGCSNNNDNCCBTGGCG 103997
 QY 360 GTGTGAGCCCTGAGTTCGGGCGCAGCGGTGACAGCGGCGGCGGCGGCGGCTCAGGT 419
 Db 103998 GCGTGGCGGCGGCSNNNDNCCGTGGCGGCTGGCGGCGGCSNNNDNCCGTGGCGGCGG 104057
 QY 420 CCGAGGCGGCTGGCGGG 437
 Db 104058 CCGGCSNNNDNCCGCGG 104075

RESULT 14
 T36481
 ID T36481 standard; cDNA; 3415 BP.
 AC T36481;
 DT 08-OCT-1996 (first entry)
 DE Human integrin beta subunit protein, beta-5, cDNA.
 KW Human; integrin beta subunit; beta-5; carcinoma; lymphoid cell;

KW immunoassay; detection; mRNA; assay; ss.
 OS Homo sapiens.
 FH Key
 FT Location/Qualifiers
 CDS 337..2736
 FT /*tag= a
 FT signal_peptide 337..405
 FT /*tag= b
 FT mat_peptide 406..2733
 FT /*tag= c
 US5527679-A.
 PD 18-JUN-1996.
 PF 01-MAY-1991; 694314.
 PR 01-MAY-1991; US-694314.
 PR 27-APR-1993; US-054077.
 PA (DAND) DANA FARBER CANCER INST INC.
 PI Hemler ME, Ramaswamy H;
 DR WPI: 96-299852/30.
 DR P-PSDB; W02194.
 PT immunoassay and mRNA hybridisation assay for beta-5 protein - useful
 PT for the detection of carcinoma(s) and to distinguish different cell
 PT types
 PS Claim 11: Columns 13-20: 21pp; English.
 CC The present sequence encodes the human integrin beta subunit
 CC protein, beta-5, which is found in carcinomas but not in lymphoid
 CC cells. An immunoassay for the detection of beta-5, comprises
 CC contacting a sample with a monoclonal antibody (Ab) which binds 1
 CC epitope of beta-5, and then with a labelled Ab which binds another
 CC epitope of beta-5, and detecting any bound label. An assay for
 CC beta-5 mRNA, comprises contacting a sample with a probe capable of
 CC hybridising to the beta-5 cDNA, and determining if binding has
 CC occurred. These assays are useful for detecting carcinomas, and for
 CC distinguishing between different cell types.
 CC Sequence 3415 BP; 752 A; 1009 G; 711 T;
 SQ Sequence 3415 BP; 752 A; 1009 G; 711 T;

Query Match 8.1%; Score 36.6; DB 1; Length 3415;
 Best Local Similarity 47.6%; Pred. No. 1;
 Matches 108; Conservative 0; Mismatches 119; Indels 0; Gaps 0;
 QY 140 ATCCAGGAGTCCCGGCTCCCTAATATTAGGCAATAGCAGCATTTGGCTGAGTGC 199
 Db 1690 AGCTGGAGTGGGGTCACTTACACTGCAGTGGCGTGGCGTGGGCTGGAACCC 1749
 QY 200 CAGAGGCTCACTCCAGGGGGACCTGGCTACTTGGCTTGGCGGCTTGGCGTCAACGGC 259
 Db 1750 AACAGCGCCAGTGCACAGGGAGCGGACCTATGCTTGGCGCTGTGTAGTGCAGCCCC 1809
 QY 260 TGCATTGTGGCTCCGCTTGGCTTGGGATGTGGCGGCGGAGACACATGTCATGTC 319
 Db 1810 GGCTACCTGGGCGGCGGCTGGAGTGGCAGGATGGGAGAACAGAGCGTGTACCAAGAC 1869
 QY 320 CAGTGGCGGCGATGAGTGGACCGGAGCGGCTGCTGTGTGTGCA 366
 Db 1870 CTGTGGCGGAGGAGCGGCAAGCCACTGTGCAGCGGCGTGGGGA 1916

RESULT 15
 V18130/c
 ID V18130 standard; DNA; 11820 BP.
 AC V18130;
 DT 04-SEP-1998 (first entry)
 DE Human chromosome 19 derived
 KW Hydronephrosis gene; HNG gene; USF2 gene; renal disease; renal aplasia;
 KW vesical-ureteral reflux; pelvi-ureteral junction obstruction;
 KW multicystic renal dysplasia; renal agenesis; hydronephrosis;
 KW Von Mayer-Rokitansky-Kuester disorder; bifid ureter; ss.
 OS Homo sapiens.
 FH Key
 FT Location/Qualifiers
 FT 5'UTR 1..1088
 FT /*tag= a
 FT CDS 1089..1189
 FT /*tag= b
 FT /product= "USF2 gene product"

/note= "contains introns"

FT exon 1089. .1150
 FT /tag= c
 FT /number= 1
 FT intron 1151. .1435
 FT /tag= d
 FT /number= 1
 FT exon 1436. .1482
 FT /tag= e
 FT intron 1483. .1571
 FT /tag= f
 FT /number= 2
 FT exon 1572. .1690
 FT /tag= g
 FT /number= 3
 FT intron 1691. .1792
 FT /tag= h
 FT /number= 3
 FT exon 1793. .1993
 FT /tag= i
 FT intron 1994. .2437
 FT /tag= j
 FT /number= 4
 FT exon 2438. .2588
 FT /tag= k
 FT intron 2589. .2708
 FT /tag= l
 FT /number= 5
 FT exon 2709. .2796
 FT /tag= m
 FT intron 2797. .3074
 FT /tag= n
 FT /number= 6
 FT exon 3075. .3132
 FT /tag= o
 FT intron 3133. .10631
 FT /tag= p
 FT /number= 7
 FT exon 10632. .10726
 FT /tag= q
 FT intron 10727. .10879
 FT /tag= r
 FT /number= 8
 FT exon 10888. .11008
 FT /tag= s
 FT intron 11009. .11099
 FT /tag= t
 FT /number= 9
 FT exon 11100. .11186
 FT /tag= u
 FT /number= 10
 FT 3'UTR 11190. .11820
 FT /tag= v
 FT polyA_signal 11723. .11728
 FT /tag= w

WO9815650-A2.

16-APR-1998.

09-OCT-1997; E05583.

09-OCT-1996; EP-202820.

(VLAAR-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.

PI Frys JFGJ, Groenen PMA, Van De Ven WJM;

WPI; 98-240833/21.

DR P-PSDB; W60569.

PT Hydronephrosis gene - useful to treat or diagnose renal diseases and disorders, e.g. vesical-ureteral reflux, pelvi-ureteral junction obstruction, multicystic renal dysplasia or renal agenesis

PS Disclosure; Fig 6A-B; 73pp; English.
 CC This DNA encodes a USF2 gene derived from human chromosome 19. A
 CC translocation partner to this gene on chromosome 6 is the hydronephrosis
 CC gene (HNG) product. The HNG gene can be used as a starting point to
 CC design suitable compounds or techniques for the treatment of renal
 CC diseases or disorders, or nucleotide probes for diagnosing cells involved
 CC in renal diseases or disorders. A protein or a fragment encoded by HNG
 CC gene can be used as a starting point for preparing suitable antibodies
 CC for diagnosing cells involved in renal diseases and disorders. The
 CC products and method can be used to treat or diagnose renal diseases and
 CC disorders selected from vesical-ureteral reflux, uni or bilateral
 CC pelvi-ureteral junction obstruction, multicystic renal dysplasia, renal
 CC agenesis, renal aplasia, Hydronephrosis, Von Mayer-Rokitansky-Kuester
 CC disorder and bifid ureter.
 SQ Sequence 11820 BP; 745 A; 1498 C; 1443 G; 683 T;

Query Match 8.1%; Score 36.6; DB 1; Length 11820;
 Best Local Similarity 47.2%; Pred. No. 1.3;
 Matches 111; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
 QY 210 CCTCCAGGGGACCTGCTACTTCCCGCCGAGGCTTCGCCCTCACCCTGCACTTGTG 269
 Db 1267 CCTCCATTTTGGAGCGGGCCCGCGCGCGTCCGGATCATGCGGCGGCGGCGGCGG 1208
 QY 270 GCTCCGCTGTGGCTCGTGGATGTGCGCGCGGAGACACATGTCTACTGCCAGTGGCGG 329
 Db 1207 CGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 1148
 QY 330 GCATGGACTGGACCGAGCGGCGCTGCTGTGTGTGTCAGCCCTGAGGTCCGGCGCAGCGG 389
 Db 1147 CGCGCGGCGAGCGGCTGCGCGGAGGATCAGACCGGGTCCAGCATGTCCATG 1088
 QY 390 TGCACAGCGCGGGCGGAGCGGCTCCAGTCCGAGGGGTTGCGGGGAGCTGA 444
 Db 1087 GGGGGCGGGGGCGGGGGCGGGGGCGGGGGCGGGGGCGGGGGAGGGAGGGGA 1033

Search completed: May 26, 2000, 03:01:51
 Job time: 10587 sec